

Isoform 1:

```

1 TTGCTCACTG CTCACCCACC TGCTGCTGCC ATGAGGCACC TTGGGGCCTT
51 CCTCTTCCTT CTGGGGGTCC TGGGGGGCCCT CACTGAGATG TGTGAAATAC
101 CAGAGATGGA CAGCCATCTG GTAGAGAAGT TGGGCCAGCA CCTCTTACCT
151 TGGATGGACC GGCTTTCCCT GGAGCACTTG AACCCAGCA TCTATGTGGG
201 CCTACGCCTC TCCAGTCTGC AGGCTGGGAC CAAGGAAGAC CTCTACCTGC
251 ACAGCCTCAA GCTTGGTTAC CAGCAGTGCC TCCTAGGGTC TGCCTTCAGC
301 GAGGATGACG GTGACTGCCA GGGCAAGCCT TCCATGGGCC AGCTGGCCCT
351 CTACCTGCTC GCTCTCAGAG CCAACTGTGA GTTTGTGAGG GGCCACAAGG
401 GGGACAGGCT GGTCTCACAG CTCAAATGGT TCCTGGAGGA TGAGAAGAGA
451 GCCATTGACA CAGCAGCCAT GGCAGGCTTG GCATTACCT GTCTGAAGCG
501 CTCAAACCTC AACCTGGTC GGAGACAACG GATCACCATG GCCATCAGAA
551 CAGTGCGAGA GGAGATCTTG AAGGCCAGA CCCCCGAGG CCACTTTGGG
601 AATGTCTACA GCACCCCAT GGCATTACAG TTCCTCATGA CTTCCCCCAT
651 GCGTGGGGCA GAACTGGGAA CAGCATGTCT CAAGGCGAGG GTTGCTTTGC
701 TGGCCAGTCT GCAGGATGGA GCCTTCCAGA ATGCTCTCAT GATTTCCTAG
751 CTGCTGCCCG TTCTGAACCA CAAGACCTAC ATTGATCTGA TCTTCCCAGA
801 CTGTCTGGCA CCACGAGTCA TGTTGGAACC AGCTGCTGAG ACCATTCCCTC
851 AGACCAAGA GATCATCAGT GTCACGCTGC AGGTGCTTAG TCTCTTGCCG
901 CCGTACAGAC AGTCCATCTC TGTTCTGGCC GGGTCCACCG TGGAAGATGT
951 CCTGAAGAAG GCCCATGAGT TAGGAGGATT CACATATGAA ACACAGGCCCT
1001 CCTTGTCAGG CCCCTACTTA ACCTCCGTGA TGGGGAAGC GGCCGGAGAA
1051 AGGGAGTTCT GGCAGCTTCT CCGAGACCCC AACACCCAC TGTGCAAGG
1101 TATTGCTGAC TACAGACCCA AGGATGGAGA AACCATTGAG CTGAGGCTGG
1151 TTAGCTGGTA GCCCTGAGC TCCCTCATCC CAGCAGCCTC GCACACTCCC
1201 TAGGCTTCTA CCCTCCCTCC TGATGTCCCT GGAACAGGAA CTCGCCTGAC
1251 CCTGCTGCCA CCTCCTGTGC ACTTTGAGCA ATGCCCCCTG GGATCACCCC
1301 AGCCACAAGC CCTTCGAGG CCCTATACCA TGGCCACCT TGGAGCAGAG
1351 AGCCAAAGCAT CTTCCCTGGG AAGTCTTTCT GGCCAAGTCT GGCCAGCCTG
1401 GCCCTGCAGG TCTCCCATGA AGGCCACCCC ATGGTCTGAT GGGCATGAAG
1451 CATCTCAGAC TCCTTGGAAC AAAACGGAGT CCGCAGGCCG CAGGTGTTGT
1501 GAAGACCACT CGTTCTGTGG TTGGGGTCCT GCAAGAAGGC CTCCTCAGCC
1551 CGGGGGCTAT GGCCCTGACC CCAGCTCTCC ACTCTGCTGT TAGAGTGGCA
1601 GCTCCGAGCT GGTGTGGCA CAGTAGCTGG GGAGACCTCA GCAGGGCTGC
1651 TCAGTGCCTG CCTCTGACAA AATTAAAGCA TTGATGGCCT GTGAAAAAAA
1701 AAAAAAAAAA AAAAAAAAAA AA

```

(SEQ ID NO:1)

#### FEATURES:

5'UTR: 1 - 30  
Start Codon: 31  
Stop Codon: 1159  
3'UTR: 1162

#### Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA 108000024653390 /altid=gi 12742775 /def=ref XP_009922.2  tr...	752	0.0
CRA 108000024636236 /altid=gi 298316 /def=gb AAB25526.1  transc...	732	0.0
CRA 18000004926133 /altid=gi 339205 /def=gb AAA61057.1  (L02648...	732	0.0
CRA 108000024042036 /altid=gi 12654675 /def=gb AAH01176.1 AAH01...	731	0.0
CRA 18000004926130 /altid=gi 4507409 /def=ref NP_000346.1  tran...	727	0.0
CRA 18000004926132 /altid=gi 339203 /def=gb AAA61056.1  (L02647...	725	0.0
CRA 18000005170902 /altid=gi 7657639 /def=ref NP_056564.1  tran...	515	e-145
CRA 18000005218941 /altid=gi 4572454 /def=gb AAD23829.1 AF12128...	501	e-140
CRA 164000136745249 /altid=gi 11968124 /def=ref NP_071979.1  tr...	481	e-134
CRA 18000004926134 /altid=gi 4507407 /def=ref NP_001053.1  tran...	108	2e-22

EST:

gi 10725490 /dataset=dbest /taxon=96...	858	0.0
gi 10947399 /dataset=dbest /taxon=96...	846	0.0
gi 9121897 /dataset=dbest /taxon=9606...	846	0.0
gi 13280819 /dataset=dbest /taxon=96...	846	0.0
gi 13287907 /dataset=dbest /taxon=96...	833	0.0
gi 13286505 /dataset=dbest /taxon=96...	831	0.0
gi 8150776 /dataset=dbest /taxon=960...	815	0.0
gi 5936410 /dataset=dbest /taxon=9606 ...	726	0.0
gi 6888875 /dataset=dbest /taxon=9606...	726	0.0
gi 6888872 /dataset=dbest /taxon=9606...	726	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

gi|10725490|adult adrenal gland  
gi|10947399| mammary gland  
gi|9121897| retinoblastoma  
gi|13280819| adenocarcinoma cell line  
gi|13287907| retinoblastoma  
gi|13286505| embryonal carcinoma, cell line  
gi|8150776| adult uterus  
gi|5936410| adult uterus  
gi|6888875| adult head\_neck  
gi|6888872| adult head\_neck

Tissue Expression:

Human leukocyte

Isoform 2:

```

1  GGAGGATTAA TCAGTGACAG GAAGCTGCGT CTCTCGGAGC GGTGACCAGC
51 TGTGGTCAGG AGAGCCTCAG CAGGGCCAGC CCCAGGAGTC TTTCCCGATT
101 CTGTGCTCACT GCTCACCCAC CTGCTGCTGC CATGAGGCAC CTTGGGGCCT
151 TCCTCTTCCT TCTGGGGGTC CTGGGGGCCC TCACTGAGAT GTGTGAAATA
201 CCAGAGATGG ACAGCCATCT GGTAGAGAAG TTGGGGCCAGC ACCTCTTACC
251 TTGGATGGAC CGGCTTTCCC TGGAGCACTT GAACCCAGC ATCTATGTGG
301 GCCTACGCCT CTCCAGTCTG CAGGCTGGGA CCAAGGAAGA CCTCTACCTG
351 CACAGCCTCA TGCTTGGTTA CCAGCAGTGC CTCCTAGGGT CTGCCTTCAG
401 CGAGGATGAC GGTGACTGCC AGGGCAAGCC TTCCATGGGC CAGCTGGCCC
451 TCTACCTGCT CGCTCTCAGA GCCAACTGGC ATGATCACAA GGGCCACCCC
501 CACACTAGCT ACTACCAGTA TGGCCTGGGC ATTCTGGCCC TGTGTCTCCA
551 CCAGAAGCGG GTCCATGACA GCGTGGTGGG CAAACTTCTG TATGCTGTGG
601 AACCTTTCCA CCAGGGCCAC CATTCTGTGG ACACAGCAGC CATGGCAGGC
651 TTGGCATTCA CCTGTCTGAA GCGCTCAAAC TTCAACCCTG GTCGGAGACA
701 ACGGATCACC ATGGCCATCA GAACAGTGGC AGAGGAGATC TTGAAGGCCC
751 AGACCCCGA GGGCCACTTT GGAATGTCT ACAGCACCCC ATTGGCATTG
801 CAGTTCTCTA TGACTTCCCC CATGCGTGGG GCAGAACTGG GAACAGCATG
851 TCTCAAGGCG AGGGTTGCTT TGCTGGCCAG TCTGCAGGAT GGAGCCTTCC
901 AGAATGCTCT CATGATTTCC CAGCTGCTGC CCGTTCTGAA CCACAAGACC
951 TACATTGATC TGATCTTCCC AGACTGTCTG GCACCACGAG TCATGTTGGA
1001 ACCAGCTGCT GAGACCATTG CTCAGACCCA AGAGATCATC AGTGTCAAGC
1051 TGCAGGTGCT TAGTCTCTTG CCGCCGTACA GACAGTCCAT CTCTGTTCTG
1101 GCCGGGTCCA CCGTGAAGA TGTCTGAAG AAGGCCCATG AGTTAGGAGG
1151 ATTACATAT GAAACACAGG CCTCCTTGTC AGGCCCTAC TTAACCTCCG
1201 TGATGGGGAA AGCGGCCGGA GAAAGGGAGT TCTGGCAGCT TCTCCGAGAC
1251 CCAAACACCC CACTGTTGCA AGGTATTGCT GACTACAGAC CCAAGGATGG
1301 AGAAACCATT GAGCTGAGGC TGGTTAGCTG GTAGCCCTG AGCTCCCTCA
1351 TCCCAGCAGC CTCGCACACT CCCTAGGCTT CTACCCTCCC TCCTGATGTC
1401 CCTGGAACAG GAACTCGCCT GACCCTGCTG CCACCTCCTG TGCACTTTGA
1451 GCAATGCCCC CTGGGATCAC CCCAGCCACA AGCCCTTCGA GGGCCCTATA
1501 CCATGGCCCA CTTGGAGCA GAGAGCCAAG CATCTTCCCT GGGAAGTCTT
1551 TCTGGCCAAG TCTGGCCAGC CTGGCCCTGC AGGTCTCCCA TGAAGGCCAC
1601 CCCATGGTCT GATGGGCATG AAGCATCTCA GACTCCTTGG CAAAAACGG
1651 AGTCCGAGG CCGCAGGTGT TGTGAAGACC ACTCGTTCTG TGGTTGGGGT
1701 CCTGCAAGAA GGCCTCCTCA GCCCAGGGGC TATGGCCCTG ACCCCAGCTC
1751 TCCACTCTGC TGTTAGAGTG GCAGCTCCGA GCTGGTTGTG GCACAGTAGC
1801 TGGGGAGACC TCAGCAGGGC TGCTCAGTGC CTGCCTCTGA CAAAATTAAA
1851 GCATTGATGG CCTGTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA

```

(SEQ ID NO:2)

#### FEATURES:

5'UTR: 1 - 131  
Start Codon: 132  
Stop Codon: 1332  
3'UTR: 1335

#### Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA 108000024636236 /altid=gi 298316 /def=gb AAB25526.1  transc...	793	0.0
CRA 108000024653390 /altid=gi 12742775 /def=ref XP_009922.2  tr...	793	0.0
CRA 18000004926133 /altid=gi 339205 /def=gb AAA61057.1  (L02648...	792	0.0
CRA 108000024042036 /altid=gi 12654675 /def=gb AAH01176.1 AAH01...	792	0.0
CRA 18000004926130 /altid=gi 4507409 /def=ref NP_000346.1  tran...	788	0.0
CRA 18000004926132 /altid=gi 339203 /def=gb AAA61056.1  (L02647...	786	0.0
CRA 18000005170902 /altid=gi 7657639 /def=ref NP_056564.1  tran...	561	e-159
CRA 164000136745249 /altid=gi 11968124 /def=ref NP_071979.1  tr...	554	e-156

FIGURE 1, page 3 of 4

CRA|18000005218941 /altid=gi|4572454 /def=gb|AAD23829.1|AF12128... 545 e-154  
CRA|18000004926134 /altid=gi|4507407 /def=ref|NP\_001053.1| tran... 128 1e-28

EST:

gi 10725490 /dataset=dbest /taxon=96...	858	0.0
gi 5936410 /dataset=dbest /taxon=9606 ...	835	0.0
gi 6888875 /dataset=dbest /taxon=9606...	726	0.0
gi 6888872 /dataset=dbest /taxon=9606...	726	0.0
gi 12258937 /dataset=dbest /taxon=960...	686	0.0
gi 10947399 /dataset=dbest /taxon=96...	680	0.0
gi 13287907 /dataset=dbest /taxon=96...	680	0.0
gi 9121897 /dataset=dbest /taxon=9606...	680	0.0
gi 13280819 /dataset=dbest /taxon=96...	680	0.0
gi 8150776 /dataset=dbest /taxon=960...	656	0.0

**EXPRESSION INFORMATION FOR MODULATORY USE:**

library source:

gi|10725490| adult adrenal gland  
gi|5936410| adult uterus  
gi|6888875| adult head\_neck  
gi|6888872| adult head\_neck  
gi|12258937| adult lung\_tumor  
gi|10947399| mammary gland  
gi|13287907| retinoblastoma  
gi|9121897| retinoblastoma  
gi|13280819| adenocarcinoma cell line  
gi|8150776|

Tissue Expression:

Human hippocampus

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted March 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Isoform 1:  
1 MRHLGAFLEFL LGVLGALTEM CEIPEMDSHL VEKLGQHLLP WMDRLSLEHL  
51 NPSIYVGLRL SSLQAGTKED LYLHSLKLG Y QQCLLGSAFS EDDGDCQGKP  
101 SMGQLALYLL ALRANCEFVR GHKGDRLV SQ LKWFLEDEKR AIDTAAMAGL  
151 AFTCLKRSNF NPGRQRITM AIRTVREEIL KAQTPEGHFG NVYSTPLALQ  
201 FLMTSPMRGA ELGTACLKAR VALLASLQDG AFQNALMISQ LLPVLNHKTY  
251 IDLIFPDCLA PRVMLEPAAE TIPQTQEIIIS VTLQVLSLLP PYRQSISVLA  
301 GSTVEDVLKK AHELGGFTYE TQASLSGPYL TSVMGKAAGE REFWQLLRDP  
351 NTPLLQGIAD YRPKDGETIE LRLVSW  
(SEQ ID NO:3)

#### FEATURES:

##### Functional domains and key regions:

PDOC00005 PS00005 PKC\_PHOSPHO\_SITE  
Protein kinase C phosphorylation site

Number of matches: 2

1	75-77	SLK
2	174-176	TVR

PDOC00006 PS00006 CK2\_PHOSPHO\_SITE  
Casein kinase II phosphorylation site

Number of matches: 6

1	67-70	TKED
2	90-93	SEDD
3	174-177	TVRE
4	226-229	SLQD
5	249-252	TYID
6	302-305	STVE

PDOC00008 PS00008 MYRISTYL

N-myristoylation site

Number of matches: 7

1	12-17	GVLGAL
2	57-62	GLRLSS
3	86-91	GSAFSE
4	149-154	GLAFTC
5	190-195	GNVYST
6	209-214	GAELGT
7	230-235	GAFQNA

PDOC00009 PS00009 AMIDATION

Amidation site

162-165	PGRR
---------	------

#### SignalP results:

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	0.37	YES
max. Y	19	0.702	0.34	YES
max. S	5	0.974	0.88	YES
mean S	1-18	0.949	0.48	YES

Most likely cleavage site between pos. 18 and 19: ALT-EM

#### BLAST Alignment to Top Hit:

>CRA|108000024636236 /altid=gi|298316 /def=gb|AAB25526.1|  
transcobalamin II, TC II [human, endothelial cells,  
Peptide, 427 aa] /org=human /taxon=9606 /dataset=nraa  
/length=427  
Length = 427

Score = 732 bits (1870), Expect = 0.0

Identities = 376/427 (88%), Positives = 376/427 (88%), Gaps = 51/427 (11%)

Frame = +1

Query: 31  
Sbjct: 1  
Query: 211  
Sbjct: 61  
Query: 391  
Sbjct: 121  
Query: 457  
Sbjct: 181  
Query: 598  
Sbjct: 241  
Query: 778  
Sbjct: 301  
Query: 958  
Sbjct: 361  
Query: 1138  
Sbjct: 421  
(SEQ ID NO:6)

Query: 31 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL 210  
MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL  
Sbjct: 1 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL 60

Query: 211 SSLQAGTKEDLYLHSLKLGYYQQCLLGSAFSEDDGDCQKGKPSMGQLALYLLALRANCEFVR 390  
SSLQAGTKEDLYLHSLKLGYYQQCLLGSAFSEDDGDCQKGKPSMGQLALYLLALRANCEFVR  
Sbjct: 61 SSLQAGTKEDLYLHSLKLGYYQQCLLGSAFSEDDGDCQKGKPSMGQLALYLLALRANCEFVR 120

Query: 391 GHKGDRLVSQLKWFLEDEKRAI----- 456  
GHKGDRLVSQLKWFLEDEKRAI  
Sbjct: 121 GHKGDRLVSQLKWFLEDEKRAIGHDHKGHPHTSYYYQYGLGILALCLHQKRVHDSVVDKLL 180

Query: 457 -----DTAAMAGLAFTCLKRSNFPNRRQRITMAIRTVREEILKAQTPEGHF 597  
DTAAMAGLAFTCLKRSNFPNRRQRITMAIRTVREEILKAQTPEGHF  
Sbjct: 181 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFPNRRQRITMAIRTVREEILKAQTPEGHF 240

Query: 598 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 777  
GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT  
Sbjct: 241 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 300

Query: 778 YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLASTVEDVLK 957  
YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLASTVEDVLK  
Sbjct: 301 YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLASTVEDVLK 360

Query: 958 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 1137  
KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI  
Sbjct: 361 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 420

Query: 1138 ELRLVSW 1158  
ELRLVSW  
Sbjct: 421 ELRLVSW 427  
(SEQ ID NO:6)

HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	829.9	8.6e-246	2
CE00052	CE00052 lymphocyte_transmembrane_protein_KAP	3.2	2.9	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11 [.]	1	11 [.]	3.2	2.9
PF01122	1/2	1	142 [.]	1	143 [.]	296.0	4.6e-85
PF01122	2/2	143	376 .]	197	450 .]	531.8	4.8e-156

FIGURE 2, page 2 of 4

Isoform 2:

```

1 MRHLGAFLEFL LGVLGALTEM CEIPEMDSHL VEKLGQHLLP WMDRLSLEHL
51 NPSIYVGLRL SSLQAGTKED LYLHSLMLGY QQCLLGSAFS EDDGDCQGKP
101 SMGQLALYLL ALRANWHDHK GHPHTSYQQY GLGILALCLH QKRVHDSVVD
151 KLLYAVEPFH QGHHSVDTAA MAGLAFTCLK RSNFNPGRRO RITMAIRTVR
201 EEILKAQTPE GHFGNVYSTP LALQFLMTSP MRGAELGTAC LKARVALLAS
251 IQDGAFQNAL MISQLLPVLN HKTYIDLIFP DCLAPRVMLE PAAETIPQTQ
301 EIISVTLQVL SLLPPYRQSI SVLAGSTVED VLKKAHELGG FTYETQASLS
351 GPYLTSVMGK AAGEREFWQL LRDPNTPLLQ GIADYRPKDG ETIELRLVSW

```

(SEQ ID NO:4)

#### FEATURES:

##### Functional domains and key regions:

PDOC00005 PS00005 PKC\_PHOSPHO\_SITE  
Protein kinase C phosphorylation site  
198-200 TVR

PDOC00006 PS00006 CK2\_PHOSPHO\_SITE  
Casein kinase II phosphorylation site  
Number of matches: 7

1	67-70	TKED
2	90-93	SEDD
3	147-150	SVVD
4	198-201	TVRE
5	250-253	SLQD
6	273-276	TYID
7	326-329	STVE

PDOC00008 PS00008 MYRISTYL  
N-myristoylation site

Number of matches: 7

1	12-17	GVLGAL
2	57-62	GLRLSS
3	86-91	GSAFSE
4	173-178	GLAFTC
5	214-219	GNVYST
6	233-238	GAELGT
7	254-259	GAFQNA

PDOC00009 PS00009 AMIDATION  
Amidation site  
186-189 PGRR

PDOC00428 PS00468 COBALAMIN\_BINDING  
Eukaryotic cobalamin-binding proteins signature  
165-178 SVDTAAMAGLAFTC

#### SignalP results:

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	0.37	YES
max. Y	19	0.702	0.34	YES
max. S	5	0.974	0.88	YES
mean S	1-18	0.949	0.48	YES

Most likely cleavage site between pos. 18 and 19: ALT-EM

#### BLAST Alignment to Top Hit:

```

>CRA|108000024636236 /altid=gi|298316 /def=gb|AAB25526.1|
transcobalamin II, TC II [human, endothelial cells,
Peptide, 427 aa] /org=human /taxon=9606 /dataset=nraa
/length=427
Length = 427

```

Score = 793 bits (2026), Expect = 0.0  
 Identities = 399/427 (93%), Positives = 399/427 (93%), Gaps = 27/427 (6%)

Query: 1 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPMDRLSLEHLNPSIYVGLRL 60  
 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPMDRLSLEHLNPSIYVGLRL  
 Sbjct: 1 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPMDRLSLEHLNPSIYVGLRL 60

Query: 61 SSLQAGTKEDLYLHSLMLGYQQCLLGSFAFSEDDGDCQKGKPSMGQLALYLLALRAN----- 115  
 SSLQAGTKEDLYLHSL LGYQQCLLGSFAFSEDDGDCQKGKPSMGQLALYLLALRAN  
 Sbjct: 61 SSLQAGTKEDLYLHSLKLGYYQQCLLGSFAFSEDDGDCQKGKPSMGQLALYLLALRANCEFVR 120

Query: 116 -----W-----HDHKGHPHTSYYYQYGLGILALCLHQKRVHDSVVDKLL 153  
 W HDHKGHPHTSYYYQYGLGILALCLHQKRVHDSVVDKLL  
 Sbjct: 121 GHKGDRLVSQLKWFLEDEKRAIGHDHKGHPHTSYYYQYGLGILALCLHQKRVHDSVVDKLL 180

Query: 154 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFPNPGRRQRITMAIRTVREEILKAQTPEGHF 213  
 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFPNPGRRQRITMAIRTVREEILKAQTPEGHF  
 Sbjct: 181 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFPNPGRRQRITMAIRTVREEILKAQTPEGHF 240

Query: 214 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 273  
 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT  
 Sbjct: 241 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 300

Query: 274 YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLAGSTVEDVLK 333  
 YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLAGSTVEDVLK  
 Sbjct: 301 YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLAGSTVEDVLK 360

Query: 334 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 393  
 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI  
 Sbjct: 361 KAHELGGFTYETQASLSGPYLTSVMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 420

Query: 394 ELRLVSW 400  
 ELRLVSW  
 Sbjct: 421 ELRLVSW 427  
 (SEQ ID NO:7)

#### HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	906.3	8.6e-269	2
CE00052	CE00052 lymphocyte_transmembrane_protein_KAP	3.2	2.9	1

#### Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11 [.]	1	11 [.]	3.2	2.9
PF01122	1/2	1	115 [.]	1	115 [.]	241.3	1.4e-68
PF01122	2/2	117	400 .]	145	450 .]	660.5	8.7e-195



1 ATATGTATGG GAAATATGCT GTCTTCCTAT TCCTACTCCC CCACCCCTCTA  
51 GCACTGAGTC CAGGTAGGTA GGCAGGGGGG TGTCTCCCTC CTTTACTTCG  
101 ACACCCCTAAC TACCTTGGGG ATCAGAAGTG ACTCTCTGGA AGGATGCTGC  
151 TGCTTCTCAC CAGAGGCTGA CGATAACGAA GGCTATCCTC CATGGCCACC  
201 TCCTCCAGGC TGCCTTCCTG GAAATAGGAA TCATAATAGT TGTACTGGA  
251 AACAGGCAGA GGGTTGGGGG AGCCAAGGCA GTCCCAACCA GGACCAAGGT  
301 GGCTCCATTG CACACACTTC ACCATGACTC CCCTGAAGGT CCAAACGTGC  
351 GGTTCGCGG AAGTTGGGCT CCCCACTGGC CTCCCTCCTT CCTCAGAACC  
401 TCCAGGGGTG CTCCTCCTAG TGGCCACATC CAGCCTTTCT GACTGGACAA  
451 CCTATCATTT AAAATTTTCA AGTAGTTCG TAAACAGACA CACGTTGCTG  
501 TATTTATTTA TGTCAGGGC TTGGTTTGTG ATAAGTCAGG CTCAAAAAGA  
551 TTGTCTTAAA AGAGTGAACC TTGGCAATTT ACCATAAAAT AATTGCAATG  
601 CAGATTGTGC ATGGAAATGA TTGGAGATAT TTTAAGGTCA TAGTGTCTTC  
651 ACAAATTGAG CTGAAAGGGA ACTGTTAGGA TGATCTTGCC TAACCCCTC  
701 ATCTCACACA GGAAGAACTA TTTTAAACTC GAGAGGTAA GTGACCTGGC  
751 CAAAGTCACA CAGCCACCAC TAGTTAACTC GTATACATTG ATTCTCCTGT  
801 GGGGCTGGGC AGATGAGGAA TCTTTTGTTC TCTCCCTGT TGCAGAGAT  
851 TTTTCTTGAG GTTACTTTCC GAGTCTGGC AAGTACCCCT GCTTCTGGTA  
901 GCTTTGTGTC CAGATTCAAT CTCATTCTTT TTATTTTATT TTATTTTGA  
951 GACAGGGTCT CACTTTGTCA CCCAAGCTGG AGTGCAGTGG TGTAATCTTG  
1001 GCTCACTGTA GCCTCCACCT CTTGGGTTCA AGCGATCCTC CTGCCTCAGC  
1051 CCCCCAAGTA GCTGGGATTA CAGACGTCTG CCACCACGCC AGGCTAATTT  
1101 ATGGTTTTTT GTATGTGTTT TTTGTGTTTT TGTAGAGACA GTGTTTCCCC  
1151 ATGTTGCCCA GGCTGGTCTC CAACTCCTGA GCTCAAGTGA TCTGCCCGCC  
1201 TCAGCCTTTC AAAGTGCTAG GATTACAGGT GTGAGCCACC GTGCCCGGAC  
1251 TTAATCCCAT TCTTTAACTT GTTTTGTGTT GTCTCTCCA GGAGGCTCCC  
1301 AGCCCTTTTC GATTGGTTGA GAAAAGTGGC CTGGCTGGTC TGGGGCCAGC  
1351 AGCACCCACC CTCCCTCAA TTGCCCAACT CCCCCCCCCA CCGAAGTGCC  
1401 CAACTCCCCC TCCCAACTG CCCAACTCCC CCACCCCCAC AATCCCTCC  
1451 CGCCACAACCT GAGGGAGGCG GTGCTGAAAA ACAGCTGACT CCAGCAATGC  
1501 TGCTCACGTG ACCACTGCAG CTGCAGCTCC CGTTCCACTC CTGTCTCTGG  
1551 GCTAGGTGGG CACTACCAGG GGCTCCTTTG GTAAGGAGTA CCGGGTAGGC  
1601 ACCCGGTCCCT GCCAATCCAC CACTGGAACA GCTGGGGGGA CAGCAGACAG  
1651 GCACGGTCGG ACAGACTTGA CAGATCAGGC ATCAGGCCCT CTGCGCTGGT  
1701 CCCGGGCTCT TTAAGCAGGA ACGTGAATGG CCTCAAGATG TCTCACATGG  
1751 TCCCACTAGC CCTCCTCCTC CTTTTGTGTC CTACCTCCAG GAGGGCTGCT  
1801 CTGCCCTTCC TTCCTCTGTT CTTTGGCCTT ATGTTCCCGG CCACCACAGG  
1851 CCTTCCCCCG CCCCACCCCT CTGCAGACTT AGCCGTGCAT TGCAGGCATG  
1901 GAGGATTAAT CAGTGACAGG AAGCTGCGTC TCTCGGAGCG GTGACCAGCT  
1951 GTGGTCAGGA GAGCCTCAGC AGGGCCAGCC CCAGGAGTCT TTCCCCGATC  
2001 TTGCTCACTG CTCACCCACC TGCTGCTGCC ATGAGGCACC TTGGGGCCTT  
2051 CCTCTTCCTT CTGGGGGTCC TGGGGGCCCT CACTGAGATG TGTGGTGAGT  
2101 AACTCGCCTC TATCCTGTGC CTCTTTCCTC CTGGGTCCCT AGTGGGGTGG  
2151 CTAGGGCATA GGATGAGGGA ACTTACCTGC CCTTCTAAGC TCCCATAGCA  
2201 GTTTGGGCTT AGCTGGACCT CAGCATTTAA CACATCCTAT TGTGATTGAT  
2251 TATATGTTTG ACTCCTCACC AGACAAGATC TCCGTTAATT CAGTCATTGC  
2301 TTCACACATT CATTCAGCGC ATACTGAGCC TTTTCTGTGT CAGGCCAGT  
2351 GTTAGCCTTT GGGGAACGTG CAAAGCATGA GACAAGTCTA ATCCCTGCCA  
2401 TCCTAGAGCT TATGTTCTAG GGAAGGGGGA CAGACAAAAG AAATGGTTAG  
2451 GTGCTCCAC CTGAAATCTC AGCATTTTGG AAGGCTGAGG CGGGAGGGGA  
2501 GGATCGCTTG AGCTCAACAG TTCAAGGTCA GCCTGGGCAA CATAGGGAGA  
2551 CCCCATCTCT ACAAAAATA AAAAAAATA AAAAAATAGT GGGCATGGGG  
2601 AAGACTTTCT GAAGACCAAG AGGACACATG GGAGCTGAAA CTCGAAGGAA  
2651 GAAAAGGAGC TGGCAGGAAA GGAGTGGGGG ACACACATTC TAGGCAGCAG  
2701 GAAGTGAGCC TTCGGAGGTC CTGCCTGCTC CAGCTCTGTG CCCCAGGGG  
2751 TCTCTTGGAG CACAGTCTCC TGGGACCTGT CTATGAGTCT GAGCTTAGAG  
2801 GCTCAGGGCT GCTCCTTCAG ACAGGAGGCA GAAGGCAGAC TTTGGGAAC  
2851 TTGGGCGGCC CACGCGCCTT TTCTCCTCCT CTGCACCTAG GATTACGTTG  
2901 AGCAATACAC TTTACCCCC ATGGTCTCTT GAGACCCTGG GGAACCCCTG  
2951 AGAGGTGGG CTAGTCATGT CCAGGTGTCA AGTGAAGAAG TCGAGGGTTG  
3001 GAGGGGCTGA GTGACCCACT CAGGGTGCTC CACCTTTTCC AGAGCTTTGC  
3051 TGAACCTAGT TTTTAGAAGT TGAAGCCTCG TTTGTTTTCG TTTGTTTTT  
3101 TGTTGAGAGA GGTTCCTCCT CTGTTGCCCA GGCTGGAGTG CAGTGGCAGG  
3151 ATCTTGCTCT ACTGAGCCT CTGCCTTGTG GGTTCAAGTG ATTTCCCCAC  
3201 CTCAGCCTCC CAAGTAGCTG GAGACTGCAT GTGCATACTA CCATGCTTGG  
3251 CTAATTTTGG TATTTTGGG TAGAGACAGG GTTTCGCCAT GTTGGCCAGG  
3301 CTGGTCTCGA ACTCCTGGGC TCAAGTGAAA CTCTTGCCCT GGCCTCCCAA  
3351 ATTGCTGAGA TTACAGGCGT GAGCCACCGT GCCCGGCCAG AACTCCAAGC  
3401 CTCTCATCTG TGTTCCATAA ATGCAATCAG ACACCTCAGG TCTGGGCCCA  
3451 GGAACCCAG CTCTTGGTTC ATGTCCGGAC AGTCCCGAGG GGAGTTCTGG  
3501 GTTCAACCAG CAAGAGCTCT TCCTCCTGGC TGATCTGGTC CTCAGCCTTG

FIGURE 3, page 1 of 22

3551 GACAGTTAGT CCATTAACCT GACCCACAG GAGCCCCAAT CCCTGGGGT  
3601 CTGGGGAATC TTGAACTGGG GTTGGGGTG CAAATATCTG CACTGAGTCA  
3651 CTTAATTGCA CCCAGCCTCA TTCCTTTATC TGTAAGTGG GCTAAGAATG  
3701 CTCCCTGCC TTCCTCTCG GTGTAGTACG AGGAAGGATC CCATGACACC  
3751 TGCTCTCCCA GTTTAAAGCT CTATATGTAT GTTGTGAAAT TGACAGGGAT  
3801 CGCTGCACAA ACGCTAATGC AAAGTGGGCT CCTGTGCTTC CTTTCTCTT  
3851 TCTTCTCTT TTTTTTTTT TTAATTTTCT TCTAGAGATG AGGTCTCACT  
3901 ATATTGCCCC GGGTTGGTTT CAAACTCCTA GGGTCAAGCG ATCCTCCAC  
3951 CTGGCCTCC CAAACTGCTG GTATTACAGG CGTGAGCCAC TCTGTCTGGC  
4001 TCCTATGCTT GTGAATGTCA ACAGCAATCA GCCCTTAGCT GGCAGGGCTG  
4051 GGTGGTAGG GCGAGAGCTC ACCCAAGGCT GCTTTTATTA CCCTGCGTGA  
4101 ATCTGCCTGG CCCCTTCTT CTAAGGAGGT TGCTCTGTGG TTGTCACTCT  
4151 CTCCTTTTAC AGCTGGATCC TGATCTTTCA GTTCTTAACC CTGTGCTGAC  
4201 TCATCGTCTG GGAAGTGAGA GCCCGGGGTG AGGTGAGGGA ACTCCCTTGC  
4251 CGGTTTCAAG AAAAGGGAAA AGGAAAGAGA GGTGAGGAGG GGGGCAGATG  
4301 ACCAGAGAGA CACAGGCTGA GAGAGACTGA GACAGACCCA GAGAGCCTCA  
4351 CACATTGAGT GACAGAGACG GAGAAATGGA GATAGGCACC AAAAATGGT  
4401 TCTCAGTGAC AGAAAGGGAA AAAAGCAACC CCCAGTCTC TCTTAACATC  
4451 TGGTGAGAAA CAGCCATGT GCTTGGTCT GGGCCACAC AGCAAGGAT  
4501 TATGTAGGGT TTCATGCTGG TGGATGGTCA CCTTATAGCA ACAGGTATCT  
4551 GGGGCTGTGG GAAAAACAGA CACGAGGTTG TGGGACCCAG ACCACAGAG  
4601 ATGGAGCTGT TCTAGGAGCT CTGGTCTCTG TTCTGGTCCC CTGGGATATG  
4651 GCACAGTGAA GGCCACCATC AGGCAGCTGG AGCCAGCAG CAACTGGGAG  
4701 GCAGTAAACA GGGACCGAAA GTGCAAGGTT ACCTCCGAGG CAAACTACTC  
4751 TAAGCTACCC TGTGCTGAGC TCAAGTCCCT TGGAACATC CTAAGGCTT  
4801 CCGCTTCCAG AGTGTTTGAG TATTTTCGTT GCACAGCTTC GAATAAATCC  
4851 CACAGCAACA GGTAAACGGC TGCAAGCTGT GACTGTTTTC TAAGAGCTCA  
4901 TCTCACAATC TCAGGTCCTC TTCATTAAAA CAGAGATGGC AGGAAAGGCG  
4951 TTATTTTGAG ATCTGCATGG AGGAAGTTCA CCAGGCAGCC TCAATTCAAC  
5001 AGCTGGAAGT TTGCGTTGTT TGGAAATTTG ATGTGTAACA CGTTCTGCAT  
5051 GTGGGCTGAT GTTTTTGTAA ACGGGTAGCA CACACATTCA GCAGGGCACC  
5101 AAAGAGCGGG GGCTTTGCAG TTAGTCCAT CCTTGGCTCT GCAGCCTTGT  
5151 GTAAGACATG ACACGACTTT GAACTTCTGT TTCTCTTCT GTGCAAGCA  
5201 ATGATGACAG TATCTACATC ACAGGACTGG CATGAGGACC AAGTGAGATT  
5251 GGGCAAGGTG CCCGGGCACA CCACTCTCAC TGTCACTGCT GATGGGCAGA  
5301 GTGGTTGCC TGGCAGTAGCA TCCTCTATCT TCAGCCACC ACCTCTCTTG  
5351 CTGGCTCACT CCAACTGCTC TTTAGAGATA CACGCTTCCC CTCTTTCTC  
5401 CTCCCACTGC CTTTCAGTAT GGCTGCATTT CCCCTGCAA GTTGGTGTGT  
5451 GCTGGGTGGA GGTGGGGTG AGGACATGTA TTCTCTGGAG AAGGCCCTGG  
5501 TAACGTCAA GCACCTTCTT GCTGGTGGCC TGGCCCTGTG ACCTCATTTG  
5551 TACCATTTTC TTTTCTAAGA AATACCAGAG ATGGACAGCC ATCTGGTAGA  
5601 GAAAGTGGGC CAGCACCTCT TACCTGGAT GGACCGGCTT TCCTGGAGC  
5651 ACTTGAACCC CAGCATCTAT GTGGGCTTAC GCCTCTCCAG TCTGCAGGCT  
5701 GGGACCAAGG AAGACCTCTA CCTGCACAGC CTCAGCTTG GTTACCAGCA  
5751 GTGCCCTCTA GGGTATTGCC AACTCTCTT TTTCCATGTC TTGCTCCACA  
5801 TACTAAGAGA TGGGAAACTT GGGTACTAGT TTGGGCTGT CACCACCTTG  
5851 TGGGCAGACC TTAGGCAAA TTTCTCCATC TATAGAATGG AGGACCTTG  
5901 TCCATCTATA GAATGAAGGG GTTGGTTGGA TTAGATCAGA GATGCTAATG  
5951 CAAGGCTCCT TTTGCTACTA CTGTCCATCA TGTGTCTGAG GCAGACATAA  
6001 CTAATCCGTG ACTATACTCT TTGATGATGA GCCCAGGAGC AGCATCTGAC  
6051 TCTATGCTCC CTTAGTGTGC CTGAGGCAGA TATCACTAAT CGATGACTGC  
6101 AGTCTTCTAC ATTGAGCTTA GAAGCAGCAT CTGACTCTGT ATGCTCCCT  
6151 CCCATGCATG AGGCAGACAT CAGTAATCCA TGACCGCATT CTTTCATACT  
6201 GAGCCAGAA GCAGCATCTT TTCTTTCTT TCCTCTCACT CTGTTGCCCA  
6251 GGCTAGAGTG CAGTGGCACA ATCTGGCTT GCCCAACCT CCAATTCCCG  
6301 GGTTCAAGTG ATTCTCGTGC CTCAGCCACC TGAATAGCTG GGAATTACAG  
6351 CGTGTGCCAC CATGCCAGC TGATTTTGT ATTTTGGTA GAGATAGGGT  
6401 TTCACCATGT TGGCCAGGCT GGTCTTGAAC TCCTGACCTC AGGTGATCCG  
6451 CCTGTCTTGG CTTCCCAAAG TGTGGGATT ATAGGCATGA GCCACTGCAC  
6501 CAATCCAAAG CAGCATCTT TGTGCTCCCT TTTCAAGAGG CATCACAGAG  
6551 AGGCCTGTTT TGGGGTTTGA ATGAGAGGCG AAGAATCAGC CATGGAGTGC  
6601 CTCTTTCTCA GACTCCCTCT TGAGAAGTGG GTGCAGGGGT GGAGAGAAAA  
6651 GAAGACTAGG CATAGTGGCT CATACTGTA ATCCCAACAT TTTGGGAGGC  
6701 TGAGGCAGGA AGATTGCTTG AGCTCAGGAG TTTGAGACCA GCCTAGGCAA  
6751 CATAGTGAGA CCACATCTCT TAAAAAAG AAAAAGAAAA AAAATGAGCC  
6801 AGGTGTAGTG ACTCATGCCT GTGGTCCCA CTCTCCGGA GGCAAGGGT  
6851 GGAGGATCTT TTGAGGCTGA GAAATCGAGG CTACAGTGAG CCATGGTGGC  
6901 ACCACTGCAC TCCAGCCTGG GAGACAGAGA GACCCTATCT CAGTAAAAAA  
6951 AAAAAATAAA AATATGGCTG GGTGTGGTGG CTCACGCCTG TAATCCCAGC  
7001 ACTTTGGGAG GCCAAGGTAG GTAGATCACA TGAGGTTAGG AGTTCGAAAC  
7051 CAGTCTGGCC AACATAGTGA AACCTGTCT CTACTGAAAA TACAAAAAT

FIGURE 3, page 2 of 22

7101 TAGCCAAGGG TGGTGGTGGG CAACTGTAAT CCCAGCTACT TGGGAGGCCG  
7151 AGGCAGAAGA ATCGCTTGAA CTCGGGAGGC GGAGGTTGCA GTGAGCTGAG  
7201 AACATGCCAC TGCACTCCAG CCTGGGCAAC AAGAGCGAAA CTCTGTCTCA  
7251 AAGAAAATAA ATAAATAAAA TAAAAAATA AAAAAGGAGG GGGCATATGG  
7301 GTGAAGTATG GACAAAAATAG TGGGGCAGGC ACAGATGATC TGGACACAGG  
7351 AGCCCTTGGA GTTTATTCTT GAATCTAACT GTTCATCTTT ATTAAATATT  
7401 TGTGGCATAA ACCTCACAAC AACATAGCCA ACACACCTCC TTTTGGAGCT  
7451 TTTATCGAAG TTTCCCACTG TTAAGATTTT TTCCCGCTTT GTGATGCGGG  
7501 TGGGGTGGGT GCTGTAAGCA GGCTTACGGG GTGGCAGTTT CTCACAAAGG  
7551 CATTAACTGG CCTTGTCCTA GGTCTGCCTT CAGCGAGGAT GACGGTGACT  
7601 GCCAGGGCAA GCCTTCCATG GGCCAGCTGG CCCTCTACCT GCTCGCTCTC  
7651 AGAGCCAACGT GTGAGTTTGT CAGGGGCCAC AAGGGGGACA GGCTGGTCTC  
7701 ACAGCTCAAA TGGTTCCTGG AGGATGAGAA GAGAGCCATT GGTGAGCAGA  
7751 CACCATCCGC TGGGGGTGGG GAGCAGCTGG GAGGGCTCAT CAGATGATAT  
7801 TCTCCAATGA GAATCAGAAC TTTGGGTTTT CTCCCAGGC GTCTTTCCCA  
7851 CCATCCATTC TGCCCATCTC ACTGCCTACG TAGAGGCTCG AACCTGTCCC  
7901 CATAGCCATC CTTGACCCAG CTTTTCCTCG GCTGCACACA TACTATTGAC  
7951 AGGTGTGTTT CGTGGTTTTT TGTTTTTTGT TTGTTTGT TTGTTGAGTT  
8001 GGAGGTTTGC TCTTGCTGCC CAGGCTGGAG TACAATGGCG CAATCTCAGC  
8051 TCACCGCAAT CTCTGCCTCC TGGGTTCAAG CAATTCTCTT GCCTCAGCCT  
8101 CCTGAGTAGC TGGGATTACA GGCATGCGCC ACCACACCCA GCTAATTTTG  
8151 TATTTTTAGT AGACGTGGGG TTTCTCCATG TTGGTCAGGC TGGTCTCGAA  
8201 CTCCTGACCT CAGGTGATCC GCTTGCCCTTA GCCTCCGAAA GTGCTGGGAT  
8251 TACAGGCATG AGCCACTGCG TTAGGCCCAC TGACAAGCCT TGTATTGGCT  
8301 AGCCACCAAG ATTGACTTGA TTATCCACCT TCGGGACAAC TGGACAGCCT  
8351 GCTTATGACT TACGCCATAG TCTGTCTCTA CTAGCTCTCC TGCCCTGACT  
8401 TGACCCAGCA TACAACAGCC AGAGCCAGCC TTTTCAATAT AAACCTGATC  
8451 TTGCTGGCAC TGCTTAAACC CTGCAGGGGC CTCGCACTGC TCCATGGCCC  
8501 AGCCTGTCTA CCCTTACCTT CTGCCCAGGC TCTGCTCATC CATTCTCTGC  
8551 CTCCACACA CCTGCCCTCT GTGGGCTCCA GCCATACCAT CTCTCAACTC  
8601 ATAAGCCAGT TTTTTCATAC AGGCTCCCTC CATCTGGACT GGCTTCCCTG  
8651 CGTGCACTT ACCTCTGCTC TACCTTTGGC TCTGCCTCCA CCCATCCTCA  
8701 GCCGTCTCCA GCATTACCTC CTTGGAGAAT CCTGCCTTGA CTTCCAGGCC  
8751 ACCCAAATAT CACTACTTGG TCTGCATTCT CGTTGCAATT GCAGTCGCAT  
8801 GAGCAATTGC TGTGGTTGAG GCCCGAAGTG CGCAAGTGCC TGTCTGCCAT  
8851 GGGTCTCTCG CTCTCTCTAA GCACAGTGCC TGACACACAG TGAGACCTCA  
8901 GCACGTATGG GCTGAGGCAA TGAAGGAATG AAGGATCCCA TGACCCAAAA  
8951 GAGCCTGTGG GAAAGTGCA GCGAGGGTCC CAGGTGCTGG CGGGGCTGGC  
9001 TGCTGGGTGG GGGCAGAGAG GCAACCCCTC TGTTTTTTTC CCTCTCAGGG  
9051 CATGATCACA AGGGCCACCC CCACACTAGC TACTACCAAT ATGGCTGGG  
9101 CATTCTGGCC CTGTGTCTCC ACCAGAAGCG GGTCCATGAC AGCGTGGTGG  
9151 ACAAACTTCT GTATGCTGTG GAACCTTTCC ACCAGGGCCA CCATTCTGTG  
9201 GGTGAGTAGG TCAGACCGTG CCAAGGCCAG GCTGGCACTC CCTCAGTCCC  
9251 CAGGTCTCTG CTGATGACGT CCATACCCCTG GCCCCACAC TCACCTTTCC  
9301 TTGGGGCTCC TCCGAATCAA GTCCCTTAGG GACGAATTGG CGAGGGCTCA  
9351 TGGGTGATGC TCCAGCTGTG AGCCAGCTTT GGAGCTGGTA GGTGGATCTC  
9401 TTGAGGCCAG GAGTTCAAGA CAACGTGGTG AAACCCCATC TCTACTAAAA  
9451 ATAAAAAAGT TAGCCGGGCA TGGTGGCACA TGCTGTAGT CCCAGCTACT  
9501 CGGGAGGCTG AGGCAGGAGA ATCACTTGAA CCTGGGAGGC GGAGGCTGCA  
9551 GTGAGTGGAG ATCGCACCCAC TGCCCTCCAG CCTGGGCAAC AGAGTGAGTG  
9601 AGACTCTGTC TCAAAAAATA AAAAAATAA TAAACTCCC CTAGTGATTG  
9651 CAATGTGCAG CTAAGTTTGG AAATAGGTGG TATGGGGTCA AGTCCCTTG  
9701 GGCCTCCCTC CTCCAGTCTT TCTCCCTAAC CTCTAGCCCT CAAGTTGCAG  
9751 AGTGATCAGC CAAACCAGTT TGCCAGAAA TGAGCAGTTT CCTGGGACAC  
9801 AGGATTTTCA GAGTCCAGAC AAGGAAAGTC TTGGGCAGAC CAGGTTGAGT  
9851 TGGTGCCTT AGCTGATCTG ACCATGTTGC CCTTCTTCTC CAAGCCCTCC  
9901 TGTGGTTGTC CATAGCTACA AGGGCCTGAC CCTCAAGCCC CTGCCTGTCC  
9951 TGGCCCTTT GGCTCTCCAG CTCATTGCAT GTTCTGTCCC CCACTTCAAG  
10001 ACACAGCAGC CATGGCAGGC TTGGCATTCA CCTGTCTGAA GCGCTCAAA  
10051 TTCAACCTTG CATGGAGACA ACGGATCACC ATGGCCATCA GAACAGTGCG  
10101 AGAGGAGATC TTGAAGGCCC AGACCCCGA GGGCCACTTT GGGAAATGTCT  
10151 ACAGACCCCC ATTGGCATTA CAGGTGGGAA AGAGACCCCTG GAGCCATGGC  
10201 CACCTTGGGG AACAGTCGGG TGGAGTGGTC AGGTGCTGGA ACACCTAGCC  
10251 CCTCCCTGCC GGCTGACCTC CTCTCTCTCT TCCTCACTCT ATCACCAGTT  
10301 CCTCATGACT TCCCCATGCT CTGGGGCAGA ACTGGGAACA GCATGTCTCA  
10351 AGGCGAGGGT TGCTTTGCTG GCCAGTCTGC AGGATGGAGC CTTCCAGAA  
10401 GCTCTCATGA TTTCCAGCT GCTGCCCTGT CTGAACCACA AGACCTACAT  
10451 TGATCTGATC TTTCCAGACT GTCTGGCACC ACGAGGTAGC CCAACTTTT  
10501 GTGGAAGCAC AGCCCTTTTAC AATCTGCTGC GCACCCATTG ACGTCCCAGT  
10551 GAGGGGAGGT TGCTTCATCC TGATTTGCTG AGTCAGCACA AGTTTGTGGG  
10601 TGTGCATGGG ACACAGTAGC CAAAATGTGG TCATAGCTTC TAGAAGCTCA

FIGURE 3, page 3 of 22

10651 CAGTGTGGGG AGGAAGACAG TAAATGGAGA TCCCTGGGCA TATCGCTTGT  
10701 GTGATACCCA GTACAGAAAT GTTTGGATGG ATGGATGGAT GGATGGATGG  
10751 ATGGATGGAT GGATGGATGG ATGAGGAGAG ACACATTTTG GTTAACTCTA  
10801 ATACAACATG ATAAGCCCCA GTAGCAGCAT GATCCAGGCT TTCTCTGAGA  
10851 GAGGGTCTGA GGACGTGACT GGGATTTGCC AATTAAGAAT GGAGAAAGAG  
10901 GCCAGGTGCA GTGACTCATG CCTGTAATCC CAACACTTTG GGAGCCCGAG  
10951 GCGGGTGGCT CACCTGAGGT CAGGAGTTCG AGACCAGCCT GGCTAACATG  
11001 GCGAAACTCC ATCTATTTAA AATACAAAAA AGTAGCTGGG TGTGGTGGCG  
11051 AGTGCCTGTA ACCCCAGCTA AGCTACTCAG GAGGCTGAGG CAAGAGAATC  
11101 ACTTGAACCT CAGAGGTGGA GGTTCAGTG AGCCAAGATC ATGCCACTGC  
11151 ACTCCAGTCT GGGTGACAGA GTAAGACTAT GTCTCAAAAA AAAAAAAAAA  
11201 AAATGGAGAA GAAGGAAGCT GGACATGGTG GCTCGTGCTT ATAATCCTAG  
11251 CACTCTGGGA AGCTGAGGCA GATGGATTGC CTGAGCCAG GAGTTTGAGA  
11301 CCAGCCTGGG CAACATGGTG AAACCCTGTC TTTACTAAAA TACGAAAGAT  
11351 TAGCCAGGCA TGGTGGTAGA CACCTATAAT CCCAGCTACT AGGGAGGCTG  
11401 AGCCACAAGA ATCACTTGAA CCTGGGAGAC AGAGGTTGCA GTGAGCCGAG  
11451 ATCGCGCCAT TGCATCCAG CCTGGGCGAC AGTGTGAGAC TCTGTCTCCA  
11501 GAAAAACAA GAATGGATAG AGTGGAGCCA AGAAGAGGCA GGAAGAACAA  
11551 AGACACAGAG GTGCACAGAG TTTGGGGGAA TTTGAGGAA TGGTCTTGCA  
11601 AAAGAGTGGG ATCTGGGAGA ATGAGTGGGA GTGGAAAGCA GATGAATGAA  
11651 GAGAAGGTGA GCGCATCAGG GTAACAGAGA TCGTGTGTA ACAAAATGCAT  
11701 GTTCTAGGAA GAGCCCTCTG GAGTGCTAGG TGCCAGAGAG GTGGGAGGAA  
11751 GGATACTGGA AGCAGAGAAA CCAGTGAGGG GCCTGATCTT GGGTGGTGGG  
11801 GAATGAGGGA CAGGGGAGGC CGGGATGGAA GCCAGGTGGT GGGGAATGAG  
11851 GGACAGGGGA GGCCGGGATG GAAGCCAGGT TTCAGCTGAG CAGGTGGCGG  
11901 TGGCATTGAT GGAGATGAGG ACATGGGGAA GGACAAAGTC CAGGTGTCTT  
11951 TGAGGGAAGA CAAGAAGACA AATAATCCAG GCTCTCTGTC CTCACACCAG  
12001 CTGCCCGCCC CTTTCTTCTT GGCACAGTCA TGTGGAACC AGCTGCTGAG  
12051 ACCATTCTCT AGACCCAAGA GATCATCAGT GTCACGCTGC AGGTGCTTAG  
12101 TCTCTTGCCG CCGTACAGAC AGTCCATCTC TGTCTGGCC GGGTCCACCG  
12151 TGGAAGATGT CCTGAAGAAG GCCCATGAGT TAGGAGGATT CACGTGAGAC  
12201 TCCCACTTCC CAGTCTCAC CCCACCAAC CTCACATGCC TGATAACAGG  
12251 GTCACAGAAA AGACGGGGAA CAGAGGAGAG GGTTCCTCG GGAGAGACAC  
12301 TGGCCCTGCT TCTGCTCTA CCTGCTCAGC TCCTTTCTTG CCCACGGTGT  
12351 TATGGAACA GGGAGCCATA GGCCAGCATT GTCATGAGA GAGCAGGCTT  
12401 TGGAGGCAGA GCCCCCAGT TGGAAATCCA ACTCTAACCA GCTAGGTTCC  
12451 AGGTAGGCAC CCACAATTCA CCGAGGAGAA CAGTTGTGCC CCTTCCCTGC  
12501 AGGGCCAGTG TGAAGAGTCC AGGAGTTAGT ACACATAGAG ATAGTGGCAT  
12551 GTGCTTTTTA TATGTGCAAG GTCCAGCACA TAGCAAGCGC TCAACACAGC  
12601 GTTGCTTTCA TCAGAGTAAG AACTGTTTTT TGTGTTGTTG TTTGTTGTTT  
12651 TTTAAGAGAC AGGGTCTCAA TCTTATCACC CAGGCTGGAG TGTAAATTGTG  
12701 CAATCACGTC TCACTGCACT CTCGAACCTT GGGGATGAAG CAACCCCTACT  
12751 GTCTTGCCTC AGCCTCCCAA ATAGCTGAGA CTATAGGCAC GTGCCACACA  
12801 ACCCTGGGTA ATTTTTTTTT TTTTTTTTTT GAGATAGGGT CTCTGTCTGT  
12851 TGCCCAGGCT GGTCTCAAA TCTTGGCCTC AAACCATCCT CACACCTGAG  
12901 GCGCTCAAAA TATTGGGATT ATAGGTGCGA GCCATCATGC TCAGCCAGAA  
12951 TAATAACTGG TTTTTTTTGT TTTTTTTTTG AGACAGAGTC TCACCTCTATT  
13001 ACCCAGGCTG TGGAGGCCCA ACTCGTGTTC GTGTATTTGT TTTTTTTTAT  
13051 TTATTTATTT ATTTTCGAGC AGAGCCTCTC TCTTTCACCT AGGCTGGAGT  
13101 GCAGTGGCGC AATCTCGGCT CACTGCAACC TCCGTCTCCT GGGTTCAGT  
13151 GATTGTCTCT CCTCAGCCTC CTGAGTAGCT GGTGCTACAG GCGCGTGCCA  
13201 CCATGCCCAG CTAATTTTGT TATTTTGTAG AGAGACAGGG TTTTACTATG  
13251 TTGGCCAGCT GGTTCCTAAC TCCTGAACCT GGGTGATCTG CCTGCCTCGG  
13301 CCTCCCAAAG TGCTGGGATT ACAGGCATGG GCCTCCGTGC CCGGCCATGT  
13351 ATTTATTTAG GCAAGGTCTC TCTCTGTTAT CCAGGCTGAA GTGCACTGGC  
13401 ACATTTCATAG CTCACTGCAG CCTCAAATTA TCCAAGTAAC AGGGACTACA  
13451 GGCATGCACC ACCACACCCA TCTACTTTTT TTTGAGATGG AGTCTCCCTC  
13501 TGTGCCCCAG ACTGGGTTGC AGTGGCACAA TTTCAGCTCA TGGCAGCATC  
13551 TACCTCCCAG GTTCAAGCGA TTCTCCTTCC TCAGTCTCCC GAGTAGCTGG  
13601 GACTATGGGC ATGACACCAC ATACCTGGCT AATGTTTATA TTTTGAGTAG  
13651 AGATGGAATT TTGCCATTTT GGCCAGGCTG GTCTTGAGCT CTTGACCTCA  
13701 AGTGATATGT CTGCCTCAGN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
13751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
13801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
13851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
13901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
13951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
14001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
14051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
14101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
14151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

FIGURE 3, page 4 of 22



17751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
17801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
17851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
17901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
17951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
18951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
19951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
20001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
20051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
20101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN  
20151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNCCAAATC  
20201 AACCGATTGC ATAATCACT CCTCTATCTT CCTTGGGGTG GAAAGTGGAT  
20251 GGGAGTTATA ATTTGAGTTC TCTTTTGCTT TAGTCCATTG AAGCTGCTAT  
20301 TACAAATATC CATAAATGCG GTGGCTTATA AACAGCAGAA ATGAGGCCGG  
20351 GTGCGGTGGC TCATGCCTAT AATTCCAGCA CTTTGGGAGG CCAAGGCAGG  
20401 TGGATCACCT GAGATCAGTA GTTCAAGACT AGCCTGACCA ACATGGTGAA  
20451 ACCCTGTCTC TACTAAAAAT ACAAAAATTT AGCTGGGGGT GGTGGCGGGC  
20501 ACCTGTAATC CCAGCTACTC AGGAGGCTGA GGCAGGAGAA TCGCTTGAAC  
20551 CCAGGAGGCG GAGGTTGCCG TGAGCTGAGA TCACGCCATT GCATTTACAGC  
20601 CTGGGCACAA AGAGTGAAAC TCCATCTCAA AATGAAATAA AATAACAGAA  
20651 ATGTATTTCT TTACAGTTCT GGAGGTTGGG TGGGCAGTCC CAGATCAGGA  
20701 CACTGACAGA TTACGTGTCT GATGGGGGCC CACTTCTGGT TGTACCTGC  
20751 TGGCTGTGTT CTCACATGGT GGAAGGAACA TGGCAACTTT CTGGGGCCTT  
20801 GTTTTTTAAT TTAAAAAAA AAAATATTTT CCTGGCCCTT GCCTGCTGAA  
20851 GGAACCTCTT TTATAATGGT ACTTAAAAAT TTTTTTTTTT GAGATGGGGG  
20901 TCTCACTCTG TCACCCACGC TGAGTGAGT ATCACAATCT CAGCTCACTG  
20951 CAACCTCTGC CTCCCTGGCT TAAGCGATCC TCCACCTCA GCCTCCTGAG  
21001 TACGTGTGAC CATAGGCCCA TGGCACAAAG CCCAGCTAAT TTTTGTATT  
21051 TTTAGTAGAA ATGTGGTTTC ACCATGTTGC ATAGGCTGGT CTCGAACCTC  
21101 TGAACCTAAG TGATCTGCCT GCCTTGGCCT CCCAAAGTGC TGGGATTCTA  
21151 GGTATGAGCC ACCCTGCTCG GCCTATAATG GCACTTTCCT ATCCCATTTGA  
21201 TGAGGCTCTA CTCTCATGAC CTAATCATCT CCCAAAGGCC CTAAGGCCTC  
21251 CTGATACCAT CACCTTTGGG GTTAGGTTTT AACATATACA TTTTGGGGGG

FIGURE 3, page 6 of 22

21301 ACACAGACAT TTTAGACCAT AGCACCTCCA TTGAAAGGAA ACATTTCTGA  
21351 CACCTGGCTA TCTCAAAGGG CCCTTTTCAGT TCCCCTGCAG GCTGCATTCC  
21401 CACATCACCA ACAAGAGCAG CGACACTCAC TCAGAGGTTA AATAACTTGT  
21451 CCAGAGTCAC AGCAGTAATG AATGACAGAG CTGGGGCTTG AATCCAGGCG  
21501 TCCTCCTAGA GCCTGGATTG TGTGTAGTGA GTGAAAGCTG ACTCCTGGGA  
21551 GACTTCTGCG TGGTCTGGT TCTCTCTCCA GACTGCACTG CGCAAGTTTC  
21601 TCTTCTGAT GGTCCCTAGG GTATTACAAA GACAGTGGCC CTGCCTGTCA  
21651 GGTGTTTTTA TTACCAGATG AGGTCTATGGC CTCAGGAACC CTGTAGGAAG  
21701 CTGAGTTCAG AGTCTTTGAG CAGGCTTTAG GGAGGTTCCA GCTTCCCACC  
21751 ACCAAGCCCC AGGTGGATTG TTACAGACTC TAGCCTCAGG GTGGGGGGTC  
21801 TGGGAAGATGA GGTTCGCGGG TCGGATATTC TGCCCAATTC GCCCTCCTT  
21851 GCTCAATCTG TTTCTGCAGG TATTGCTGAC TACAGACCCA AGGATGGAGA  
21901 AACCATTGAG CTGAGGCTGG TTAGCTGGTA GCCCTGAGC TCCCTCATCC  
21951 CAGCAGCCTC GCACACTCCC TAGGCTTCTA CCCTCCCTCC TGATGTCCCT  
22001 GGAACAGGAA CTCGCCTGAC CCTGCTGCCA CCTCCTGTGC ACTTTAGACA  
22051 ATGCCCCCTG GGATCACCCC AGCCACAAGC CCTTCGAGGG CCCTATACCA  
22101 TGGCCCACTT TGGAGCAGAG AGCCAAGCAT CTTCCTGGG AAGTCTTTCT  
22151 GGCCAAGTCT GGCCAGCCTG GCCCTGCAGG TCTCCCATGA AGGCCACCCC  
22201 ATGGTCTGAT GGGCATGAAG CATCTCAGAC TCCTTGCCAA AAAACGGAGT  
22251 CCGCAGGCCG CAGGTGTTGT GAAGACCACT CGTTCTGTGG TTGGGGTCTT  
22301 GCAAGAAGGC CTCCTCAGCC CGGGGGCTAT GGCCCTGACC CCAGCTCTCC  
22351 ACTCTGCTGT TAGAGTGGCA GCTCCGAGCT GGTGTGGCA CAGTAGCTGG  
22401 GGAGACCTCA GCAGGGCTGC TCAGTGCCTG CCTCTGACAA AATTAAAGCA  
22451 TTGATGGCCT GTGGACCTGC TACAGTGGCC TGGTGCCTCA TACTCCTCAG  
22501 GTGAGGGGGC AGGGACAAGA GAAGGGGGAA GTAACCCCAT CAGGGAGGAG  
22551 TGGAGGGTGC CTGAGCCGCC ATGTGGGCAT TGGGGGAGTG ATGGGAATGC  
22601 CAGCAGTGAT GACGTTGACT ACTGACTGAG CACCCACTAC TATGACTGAG  
22651 CACTCACTCG CTAGATACTA TCTTGAACATG CTCTGTGAGG TTGTTGATAT  
22701 TTTCAATTTT ATCTGTGCTT TACAAATCAG GAAACTGGGA GGCCCGGGCGT  
22751 GGTGGCTCAC GCCTGTAATC CCAGCACTTT AGGAGGCCAA GGCAGGTGGA  
22801 TCACAAGGTC AGGAGTTTGA GATCAGCCTG GCCAACATGG TGAAACTCCA  
22851 TCTTTACTAA AAATACAAAA AATTAGCCAG GCATGGTGTG GCATGCCTGC  
22901 ATGCCTGTAA TCCAGTTTAC TTGGGAAGCT GAGGCAGGAG AATTGCTTGA  
22951 ACCCTGGAGG CGGAGGTTGT AGTGAGCCGA GATCACGCCA TTGCACTCCA  
23001 GCTTGGGCAG GAAGAGAAAC ACTCTCAAAA AAAAAAATAA ATCAGGAAAC  
23051 TGGTGCTCAA AAAGGAAAAA TGACTCACCA AGGTCACAGA CTAGGCAGTG  
23101 ATGCTGGGGG AACCTGGCTC AGGGGACACA GACCTGGCCT GGGGCAGCCT  
23151 TGCAGCTCCT CCACTAAAAA ACTGAAAATG AGGGGCTTCG ATGATGGTTA  
23201 TAATCGTATG GCAGAGCCCC AACTCAACTG GAGCCCTGGG ACCCAGAAGC  
23251 TAGGCTCTCA CTCCTGCTT TTCCACAAGG CACCATTAGG GCATCACCCC  
23301 AGGCCTCGGC AGCCACGACG CAGGGATCCT GCCTCTCATT GGTGGGGGGC  
23351 TTAGGGGCTC TGGGCTGCCC TCTTGAAGAG GGGGTTTACG CCAGCGAGGC  
23401 ACCCCCTATG CTGCACCCCA CCAAGGTTAG GAAGAGGTCC TGTCTCAGT  
23451 GGGGCCCTCT GATGAACAGC CCATCAGGTC TGCGTCCACA TGCTTGGAA  
23501 GAGATGGTGA CATACTCAA GTCTTGAAG CCGCATATTA AACCACCTAG  
23551 AGCACCATCT TCAAAACATTT AGGGTCTGAG AAGATAGGGG AAGTAAGCAA  
23601 TTTAAACAT TTTCTTATAT TGGGCCAGST GCAATGGCTC ACGTCTGTAA  
23651 TCCCAGCGCT TTGGGAGGAC GAGGATCACC TGAGGTGAGG AGTTCAAGAT  
23701 CAGCCTGGCC AACATGGAGA AACCCCATCT CTAATAAAAA TACAAAAATT  
23751 AGCTCAGGCG TGGTGATGTG CACCTGTAAT CCTAGCTATT CAGGAGGCTG  
23801 AGGCACAAGA ATTGCTTGAG TCAATATTGC ACCACTGCAC TCCAGCCTGG  
23851 GCAACAGCGA GACTCTTGTC TCAAAAAAAA AAAAAGATAT TTGCTGAAAA  
23901 GACCCAGCCT GCCAAACTCA GGGGCAGCCA AGGGAGGTAG TGAAATGGAA  
23951 GTTGAGGCTC AGCGCTCCCA CACCTCCACT GCCCTCAGGC CTCTCTGCCC  
24001 TCTTTCCCAT CAGTCAGCTG CTTCTGGGCA TGGTCTGCG AGAGACTTGG  
24051 CCTCCTTCCA GTTCAAGCTC CCTCTTAGAT TGTGTCCAC GCCACTGAGT  
24101 CTTTGGGACA CTGGGTGAGA TGTCTAGTCT GGCACAATTG GCAGGAATCC  
24151 CAAGAAACAG TGTGAGTGAG GGGACAGTCG TGTGAGTGC CCTCCATCTG  
24201 GGACTGGGAG GCAGGTCTAT GTCAGGCCTG CATTTAGATC TCTAATGGCT  
24251 CCAGACAAGC CCCTTCAGCT CACTAAGCCT GTTTCCTAAC ACAGCTGTGG  
24301 GATGGTGCTT TGGTTTACAT AGCACGCGAT ACCATCATAG ATCACATGGG  
24351 GAAACTGAGG CCCAGGAGT GATCTGCTGG CACATGCAGT GACAAGAGGA  
24401 GAGGCCCATC TCAGCCTTGC AGCAAGGTTG CCAGAAATCG ATTCTCGCCC  
24451 CCATCCCGTA AAGATAGCTG GGATTACAGG TGTGCACCAC CATGCCAGC  
24501 CTAATTTTTG TATTATTAGT AGAGATGGGG TTTCACCATG TTGTCCAGGC  
24551 TGGTCATGAA CTCCTGACCT CAAGTGATCC ACCCGCTTTG GCCTCCCAAA  
24601 GTGCTGGGAT TACAAGCATG AGCCACAGTG CCTGGCCTGA CCCTGCTCTT  
24651 TTGAAAGACC ATTCCCCCAA ATCTGTGCA CCTGTGTGCC TTTCTTCTCT  
24701 CTGCCTCCTC TCAGCTCTGC CCCGCTCTCC TCCCTTCTCC TCTGGCAAA  
24751 CCCACTCATC TCTTGAAGCC CTTCTTCCAG GGAAGCCCT GATCATGCTG  
24801 CTTTCTCCTG TGGGAGGGAT GAAGGACGTG GCCACGGAG TTTGTTTTGT

FIGURE 3, page 7 of 22

```

24851 TTTGTTTTGA GATGGAGTTT TGCTCATGTT GCCCAGGCTG GGGTACAATG
24901 GTACGATCTC AGCTCACTGC AACCTCTACG TCCCGGGTTC AAGCGGTTCT
24951 CCTGCCCTTAG CCTCCCCAGT AGCTGGGATT ACTGGCATGA ACCACCACAC
25001 CTGGCTAATT TTGTGTTTTT AGTAGAGATG GGGTTTCTTC ATGTTGGTCA
25051 GGCTGGTCTC GAACTCCCAA CCTCAGGTGA TCTGCCTGCC TCGGCCTCCC
25101 AAAGTACTGG GATTACAGGG TTGAGCCACT GTGCCTGGCC CAGGCCACAG
25151 GAGTTTTAAG AGGCTTCCTG TGGCAGTGGC ATCCAGACGG AGTGCAGAAA
25201 CTCAAAGTTG AAGGCCAGAA GCTCAGGGAA GGGGGAGTGT GAGTTGAGGA
25251 GTCTCTTGGC TGCCAGGGCC AGAAACCGAA CTCCAAGCCT CTCCACAACA
25301 GCGGGTGTAG AGCATGTAGA ATCAGAGAGG AGGCTGAGCC ATGCAGCCCC
25351 GAGAAGAGGG GAATGCCACT GAGCCACAGA GACCCAGTGC CACTGCCAGG
25401 TGTCTCTGCC TCCACTTCCC ATGACCCGGC CTGTCTCTGT ATGCAGGCTT
25451 CACCCTCTCT CGTTGTACAT TGTACACATT CTAGGTGACA CCAGCAGCTT
25501 CTGATTCTCA TCTCCCATAA CATCAGCCCC CCAGAGAGGG GACAACTGCT
25551 GAGCTGTATA CATAATAGAT GCCCCTTTCC TGGAGGCCAT GGTATGGTC
25601 AGCGTGGAGA GGATGAAGCC TGAGCAGGCA GGATCGGGGG TCTAGAGGGG
25651 AAGGAGGTGG AAGTTGAGAT CACAGACCTG TGGTCAGGTG GCCTGGGAAG
25701 GGTTTGACGA GTGTCGGCCC AAAGAGCTTG GAAGGGATTG TGCTGCTGTG
25751 GGTGAGCACT GCCTCTCCCC TTAGGGACAA CAGCCACCTC TTCTCTCCCC
25801 ATTTGCCCTT CCCTTCTGTA GATATGAAAC ACAGGCCTCC TTGTCAGGCC
25851 CCTACTTAAC TCCCGTGATG GGGAAAGCGG CCGGAGAAAG GGAGTTCTGG
25901 CAGCTTCTCC GAGACCCCAA CACCCCACTG TTGCAAGGTG AGTCATGGCC
25951 TGACACTCTG GATGTGTCCC CTACCCCAAG CTTACTCAGC CAAGAGGCTT
26001 CATCAACTCA CCCCAGCTTT CCCTAGCACC CTCCTGGGCC ACACCTTCAC
26051 AAAATCACTG ATGCTCAAAG TTGGATATAA TATATTGAAC TGAAGCCTTA
26101 GCATTTTAT GCAAGTACT GTGGAAATTC TAGGAAACCA GACAGATTAC
26151 AAAAAAAAAA AAAAACTAGA AGAAAATTAA CATCACCTAG GATATACTAC
26201 CTAGGAATAA CGTCTTTTAT TTTGAGATGG AGTTTCGCTC TTGTTGCCCA
26251 GGCTGGAGTG CAGCGGTATG ATCTCGGCTC GCTGCAACCT CCGCCTCCTG
26301 GGTTCATGTG ATTCTTCCAC CTCGGCCTTC CTAGAGCCCA AGTGGTCTGC
26351 CTGCCTCTGC CTCCCAAAGT TCTGGGATTA CAGGCATGAG CCACCGCACC
26401 CAGCCAAAAT TACTTAACTT TTCTTCTAGA TACTTTTTAA AAATATGGCA
26451 GTAAGTTTTT CATAAAAAAT GGAGCCATGC TATCCAGTGG AAATTTAATG
26501 TTGCCACAT GTATAACTTA AAAATTTTCA ATATGTGTAT ACATATATAT
26551 GAAATATATA TATACAGACA CACATATATA TGTATACATA TATATACACA
26601 TATATATGTA TACATATATA CACACATATA TGTATACATA TATATACACA
26651 CATATACACA TATATACACA CACATACATA TATACACACA CATATATACA
26701 CACATATATA CACACATGCA CACATATATA TGTATACATA TATACACACA
26751 TGTATACGTA TATATACACA CATATATACA CACATATATA TACACACATA
26801 TACACACATA CACACACATA TATACACACA TATACACACA CATATATACA
26851 CACATATATA TGTATACATA TATATACACA CATATATACA CACATACACA
26901 TACATATATA CACATATACA CATATACACA CACATATACA CACATGTATA
26951 CATATATATA CACACATGTA TACATATGTA TACACACACA TATATGTATA
27001 CATATATACA CACATACATA TGTGTACATA TATACACACA TACATATGTA
27051 TACATATATA CACACAT

```

(SEQ ID NO:5)



Isoform 1:

**FEATURES:**

Exon: 2031-2094  
 Intron: 2095-5569  
 Exon: 5570-5762  
 Intron: 5763-7571  
 Exon: 7572-7741  
 Intron: 7742-10000  
 Exon: 10001-10173  
 Intron: 10174-10298  
 Exon: 10299-10485  
 Intron: 10486-12027  
 Exon: 12028-12193  
 Intron: 12194-25821  
 Exon: 25822-25939

**Allelic Variants (SNPs):**

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
921	C	T	Beyond ORF(5')			
1781	C	T	Beyond ORF(5')			
1850	G	A	Beyond ORF(5')			
2839	A	G	Intron			
3730	G	A	Intron			
6631	G	A	Intron			
6945	-	A	Intron			
6952	A	T	Intron			
7457	G	A	Intron			
7830	T	A	Intron			
8089	T	C	Intron			
8551	C	T	Intron			
9269	G	C	Intron			
9362	C	T	Intron			
9782	G	T	Intron			
11493	G	A T	Intron			
12260	A	G	Intron			
13086	T	C	Intron			
13183	T	C	Intron			
21240	C	G	Intron			
21695	A	G	Intron			
22058	C	T	Intron			
22233	C	A	Intron			
22245	C	-	Intron			
22375	C	T	Intron			
23042	A	- T	Intron			
23344	T	C	Intron			
23873	A	-	Intron			
24764	G	T	Intron			
24939	T	C	Intron			
24945	G	A	Intron			
25092	C	T	Intron			
25428	T	G	Intron			
25513	C	T	Intron			
25684	C	T	Intron			
26165	A	-	Beyond ORF(3')			

Context:

DNA

Position

921 TTGGAGATATTTTAAGGTCATAGTGTCTTCACAAATTGAGCTGAAAGGGAAGTGTAGGA  
 TGATCTTGCCTAACCCCTCTCATCTCACACAGGAAGAAGTATTTTAACTCGAGAGGTAA  
 GTGACCTGGCCAAAGTCACACAGCCACCACTAGTTAACTCGTATACATTGATTCTCCTGT  
 GGGGCTGGGCAGATGAGGAATCTTTGTTCTCTCCCTGTTGCAGAGATTTTTTTGAG  
 GTTACTTTCCGAGTTCTGGCAAGTACCCCTGCTTCTGGTAGCTTTGTGTCTCGATTCAAT  
 [C,T]  
 TCATTCTTTTATTTTATTTTATTTTGGAGACAGGGTCTCACTTTGTACCCCAAGCTGGA  
 GTGCAGTGGTGTAATCTTGGCTCACTGTAGCCTCCACCTCTTGGGTTCAAGCGATCCTCC

TGCTTCAGCCCCCAAGTAGCTGGGATTACAGACGTCTGCCACCACGCCAGGCTAATTTA  
TGGTTTTTTGTATGTGTTTTTTGTGTTTTTTGTAGAGACAGTGTTCCTCCCATGTGCCCAG  
GCTGCTCTCAAACCTCTGAGCTCAAGTGATCTGCCCGCCTCAGCCTTTCAAAGTGCTAGG

1781 ACAGCTGACTCCAGCAATGCTGCTCACGTGACCACTGCAGCTGCAGCTCCCGTTCCACTC  
CTTGTCTTGGGCTAGGTGGGCACTACCAGGGGCTCCTTTGGTAAGGAGTACCGGGTAGGC  
ACCCGGTCTCTGCAATCCACCCTGGAACAGCTGGGGGGACAGCAGACAGGCACGGTCGG  
ACAGACTTGACAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCGGGCTCTTTAAGCAGGA  
ACGTGAATGGCTCAAGATGCTCACATGGTCCCCTAGCCCTCCTCCTCCCTTTGTTC  
[C, T]  
TACCTCCAGGAGGGTCTCTGCCCTTCTCTCTGTCTTTGGCCTTATGTTCCTCCG  
CACCACAGGCCTTCCCCCGCCCCACCCCTCTGCAGACTTAGCCGTGCATTGCAGGCATGG  
AGGATTAATCAGTGACAGGAAGCTGCGTCTCTCGGAGCGGTGACCAGCTGTGGTCAGGAG  
AGCCTCAGCAGGGCCAGCCCCAGGAGTCTTTCCCGATTCTTGCTCACTGCTCACCACCT  
GCTGCTGCCATGAGGCACCTTGGGGCCTTCTCTCTCTTCTGGGGTCTTGGGGGCCCTC

1850 GGCTAGGTGGGCACTACCAGGGGCTCCTTTGGTAAGGAGTACCGGGTAGGCACCCGGTCC  
TGCCAATCCACCCTGGAACAGCTGGGGGGACAGCAGACAGGCACGGTCGGACAGACTTG  
ACAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCGGGCTCTTTAAGCAGGAACGTGAATG  
GCCTCAAGATGCTCTCACATGGTCCCCTAGCCCTCCTCCTCCCTTTGTTCCTACCTCCA  
GGAGGCTGCTCTGCCCTTCTCTCTGTCTTTGGCCTTATGTTCCTCCGCCACCACAG  
[G, A]  
CCTTCCCCCGCCCCACCCCTCTGCAGACTTAGCCGTGCATTGCAGGCATGGAGGATTAAT  
CAGTGACAGGAAGCTGCGTCTCTCGGAGCGGTGACCAGCTGTGGTCAGGAGAGCCTCAGC  
AGGGCCAGCCCCAGGAGTCTTTCCCGATTCTTGCTCACTGCTCACCACCTGCTGCTGCC  
ATGAGGCACCTTGGGGCCTTCTCTCTTCTGGGGTCTTGGGGGCCCTCACTGAGATG  
TGTTGTGAGTAACTCGCCTCTATCCTGTGCCTCTTTCTCTGGGTCTTAGTGGGGTGG

2839 AACATAGGGAGACCCCATCTCTACAAAAATAAAAAAATAAAAAATAGCTGGGCATGG  
GGAAGACTTTCTGAAGACCAAGAGGACACATGGGAGCTGAAACTCGAAGGAAGAAAAGGA  
GCTGGCAGGAAGAGGAGTGGGGGACACACATTCTAGGCAGCAGGAAGTGAGCCTTCGGAGG  
TCCTGCCTGCTCCAGCTCTGTGCCCCAAGGGTCTCTTGGAGCACAGTCTCCTGGGACCT  
GTCTATGAGTCTGAGCTTAGAGGCTCAGGGCTGCTCCTTCAGACAGGAGGCAGAAGGCAG  
[A, G]  
CTTTGGGAACCTTTGGGCCGCCACGCGCCTTTTCTCCTCCTCTGCACCTAGGATTACGTT  
GAGCAATACACTTTCACCCCATGGTCTCTTGGAGCCCTGGGGAACCCCTGAGAGGTGGG  
TGCAGTCATGTCCAGGTGTCAAGTGAAGAAGTCGAGGGTGGAGGGGCTGAGTGACCCAC  
TCAGGGTGTCTCCACCTTTTCCAGAGCTTGTCTGAACCTAGTCTTTTAGAACTTGAAGCCTC  
GTTTGTTCGTTTTGTTTTTTGTTGAGAGAGGTTCTCCCTCTGTTGCCAGGCTGGAGT

3730 GACACCTCAGGTCTGGGCCAGGAACCCAGCTCTTGGTTCATGTCCGGACAGTCCCCAG  
GGGAGTTCTGGGTTCACCCAGCAAGAGCTCTTCTCCTGGCTGATCTGGTCTCAGCCTT  
GGACAGTTAGTCCATTAACTGACCCACAGGAGCCCCAATCCCTTGGGGTCTGGGGAAT  
CTTGAACCTGGGGTTTGGGTGCAAAATATCTGCACTGAGTCACTTAATTGCAACCCAGCCTC  
ATTCTTTATCTGTAAAGTGGGCTAAGAATGCTCCCTGCCTTCTCCTCGGTGTAGTAC  
[G, A]  
AGGAAGGATCCCATGACACCTGCTCTCCAGTTTAAAGCTCTATATGTATGTTGTGAAAT  
TGACAGGGATCGCTGCACAAACGCTAATGCAAAGTGGGCTCCTGTGCTTCTCTTTCTCTT  
TCTTCTTCTTTTTTTTTTTTTTAATTTCTTCTAGAGATGAGGTCTCACTATATTGCCCA  
GGGTGTTGTTTCAAACTCCTAGGTCAAGCGATCCTCCACCTTGGCCTCCCAAACCTGCTG  
GTATTACAGGCGTGAGCCACTCTGTCTGGCTCCTATGCTTGTGAATGTCAACAGCAATCA

6631 TGAATAGCTGGGATTACAGGCGTGTGCCACCATGCCAGCTGATTTTTGTATTTTTTGGTA  
GAGATAGGGTTTACCCTGTTGGCCAGGCTGGTCTTGAACCTCTGACCTCAGGTGATCCG  
CCTGCTTGGCTTCCCAAAGTGTGGGATTATAGGCATGAGCCACTGCACCAATCCAAAA  
GCAGCATCTTTGTGCTCCCTTTCAAGAGGCATCACAGAGAGGCTGTTTTGGGGTTTGA  
ATGAGAGGCGAAGAAATCAGCCATGGAGTGCCTCTTCTCAGACTCCCTCTTGAGAAGTGG  
[G, A]  
TGACAGGGGTGGAGAGAAAAGAACTAGGCATAGTGGCTCATACCTGTAATCCCAACATT  
TTGGGAGGCTGAGGCAGGAAGATTGCTTGAAGCTCAGGAGTTTGAGACCAGCCTAGGCAAC  
ATAGTGAGACCACATCTCTTAAAAAAAAGAAAAAGAAAAAATGAGCCAGGTGTAGTGA  
CTCATGCTGTGGTCCCACTTCTCCGAGGCAAGGTGGGAGGATCTTTTGGGCTGAG  
AAATCAGGGCTACAGTGAAGCATGGTGGCACCCTGCACTCCAGCCTGGGAGACAGAGAG

6945 AGAAAAGAAGACTAGGCATAGTGGCTCATACCTGTAATCCCAACATTTTGGGAGGCTGAG  
GCAGGAAGATTGCTTGAGCTCAGGAGTTTGAGACCAGCCTAGGCAACATAGTGAGACCAC  
ATCTCTTAAAAAAAAGAAAAAGAAAAAATGAGCCAGGTGTAGTGACTCATGCTGTGG  
TCCCACTTCTCCGAGGCAAGGTGGGAGGATCTTTGAGGCTGAGAAATCAGGCTAC  
AGTGAGCCATGGTGGCACCCTGCACTCCAGCCTGGGAGACAGAGACCCTATCTCAGT  
[-, A]  
AAAAAAAATAAAAAATATGGCTGGGTGTGGTGGCTCACGCCTGTAATCCCAAGCACTTT

FIGURE 3, page 10 of 22

Variable	Mean	SD	Min	Max	Skewness	Kurtosis	Normality
Age	35.5	10.5	20	65	0.1	3.0	0.95
Gender	1.5	0.5	1	2	0.0	3.0	0.95
Marital Status	1.5	0.5	1	2	0.0	3.0	0.95
Education	12.5	2.5	9	16	0.1	3.0	0.95
Income	1500	500	500	3000	0.2	3.0	0.95
Occupation	1.5	0.5	1	2	0.0	3.0	0.95
Health Status	1.5	0.5	1	2	0.0	3.0	0.95
Stress Level	2.5	1.0	1	4	0.1	3.0	0.95
Life Satisfaction	3.5	1.0	1	5	0.1	3.0	0.95
Resilience	3.5	1.0	1	5	0.1	3.0	0.95
Optimism	3.5	1.0	1	5	0.1	3.0	0.95
Emotional Stability	3.5	1.0	1	5	0.1	3.0	0.95
Self-Esteem	3.5	1.0	1	5	0.1	3.0	0.95
Life Satisfaction	3.5	1.0	1	5	0.1	3.0	0.95
Resilience	3.5	1.0	1	5	0.1	3.0	0.95
Optimism	3.5	1.0	1	5	0.1	3.0	0.95
Emotional Stability	3.5	1.0	1	5	0.1	3.0	0.95
Self-Esteem	3.5	1.0	1	5	0.1	3.0	0.95

9269 AGGCCAGGGTCCCAGGTGCTGGCGGGGCTGGCTGCTGGTGCGGGGAGAGAGGCCAACCC  
TCGTGTTTTTTCCCTCTCAGGGCATGATCACAAAGGGCCACCCCACTAGCTACTACCA  
GTATGGCTGGGCATTCTGGCCCTGTGTCCTCACCAGAAGCGGGTCCATGACAGCGTGGT  
GGACAACTTCTGTATGCTGTGGAACCTTCCACAGGGCCACCATTCGTGGTGAGTA  
GGTCAGACCGTGCCAAGGCCAGGCTGGCACTCCCTCAGTCCCCAGGTCTGCATGATGAC  
[G,C]

TCCATACCCCTGGCCCCACACTCACCTTTTCCTTGGGGCTCCTCCGAATCAAGTCCTTTAG  
GGACGAATTGGCGAGGGCTCATGGGTGATGCTCCAGCTGTGAGCCAGCTTTGGAGCTGGT  
AGGTGGATCTCTTGGAGCCAGGAGTTCAAGACAACGTGGTGAACCCCATCTCTACTAAA  
AATAAAAAAGTTAGCCGGGCATGGTGGCACATGCCTGTAGTCCAGCTACTCGGGAGGCT  
GAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGCTGCAGTGAGTGGAGATCGCACCA

9362 GGGCCACCCCCACACTAGCTACTACCAGTATGGCCTGGGCATTCTGGCCCTGTGTCTCCA  
CCAGAAGCGGGTCCATGACAGCGTGGTGGACAACTTCTGTATGCTGTGGAACCTTTCCA  
CCAGGGCCACCATTTCTGTGGGTGAGTAGGTGAGACCGTGCCAGGGCCAGGCTGGCACTCC  
CTCAGTCCCCAGGTCTGCATGATGACGTCCATACCCTGGCCCCACACTCACCTTTTCCT  
TGGGGCTCCTCCGAATCAAGTCTTTAGGGACGAATTGGCGAGGGCTCATGGGTGATGCT  
[C, T]  
CAGCTGTGAGCCAGCTTTGGAGCTGGTAGGTGGATCTCTTGGAGCCAGGAGTTCAAGACA  
ACGTGGTGAACCCCATCTCTACTAAAAATAAAAAAGTTAGCCGGGCATGGTGGCACATG  
CCTGTAGTCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGG  
AGGCTGCAGTGAGTGGAGATCGCACCACTGCCCTCCAGCCTGGGCAACAGAGTGAGTGAG  
ACTCTGTCTCAAAAAATAAAAAATAAAATAAACTCCCTAGTGATTCCAATGTGCAGCT

9782 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCG  
GAGGCTGCAGTGAGTGGAGATCGCACCACTGCCCTCCAGCCTGGGCAACAGAGTGAGTGA  
GACTCTGTCTCAAAAAATAAAAAATAAACTCCCTAGTGATTCCAATGTGCAGC  
TAAGTTTGAAATAGGTGGTATGGGGTCAAGTCTCTTGGGCTCCCTCCTCCAGTCCCT  
CTCCCTAACCTCTAGCCCTCAAGTTGCAGAGTGATCAGCCAACCAAGTTTGGCCAGAAAT  
[G, T]  
AGCAGTTTCTGGGACACAGGATTTTCAGAGTCCAGACAAGGAAAGTCTTGGGCAGACCA  
GGTTGAGTTGGTGCCCTTAGCTGATCTGACCATGTTGCCCTTCTTCCAGCCCTCCTG  
TGGTTGTCCATAGCTACAAGGGCTGACCCCTCAAGCCCTGCCTGTCTGGCCCCCTTGG  
CTCTCCAGCTCATTCATGTTCTGTCCCCACTTCAAGACACAGCAGCCATGGCAGGCTT  
GGCATTCACTGTCTGAAGCGCTCAAACTTCAACCTGGTGGAGACAACGGATCACCAT

11493 AAAAAAAAAAATGGAGAAGAAGGAGCTGGACATGGTGGCTCGTGCTTATAATCCTAGCA  
CTCTGGGAAGCTGAGGCAGATGGATTGCCTGAGCCAGGAGTTTGAGACCAGCCTGGGCA  
ACATGGTGAACCCCTGTCTTTACTAAAAACGAAAGATTAGCCAGGCATGGTGGTAGACA  
CCTATAATCCCAGCTACTAGGGAGGCTGAGGCCAAGAATCACTTGAACCTGGGAGACAG  
AGGTTGCAGTGAGCCGAGATCGCGCCATTGCACTCCAGCCTGGGCGACAGTGAGACTC  
[G, A, T]  
GTCTCCAGAAAAAACAAGATGGATAGAGTGGAGCCAAGAAGAGGCAGGAAGAACAAGA  
CACAGAGGTGCACAGAGTTTGGGGGAATTTTGAGGAATGGTCTTGCAAAAGAGTGGGATC  
TGGGAGAATGAGTGGGAGTGGAAAGCAGATGAATGAAGAGAAGGTGAGCGCATCAGGGTA  
ACAGAGATGCGTTGTGAACAAATGCATGTTCTAGGAAGAGCCCTCTGGAGTGCTAGGTGC  
CAGAGAGTGGGAGGAAGGATACTGGAAGCAGAGAAACCAGTGAGGGGCTGATCTTGGG

12260 ACAAGAAGACAAATAATCCAGGCTCTCTGTCTCACACCAGCTGCCCGCCCTTTCTTCC  
TGGCACAGTCACTGTTGGAACCACTGCTGAGACCATTCCTCAGACCCAAGAGATCATCAG  
TGTCACGCTGCAGGTGCTTAGTCTCTTGCCGCGGTACAGACAGTCCATCTCTGTTCTGGC  
CGGGTCCACCGTGGAAAGATGCTCTGAAGAAGGCCATGAGTTAGGAGGATTACAGTGAGA  
TCCCCACCTCCAGTCTCACCACCCCAACCTCACATGCCTGATAACAGGGTCACAGAA  
[A, G]  
AGACGGGGAACAGAGGAGAGGGTTCCCTCGGGAGAGACACTGGCCCTGCTTCTGCTTCTA  
CCTGCTCAGCTCCTTTCTTGCCACGGTGTATGGAACAGGGAGCCATAGGCCAGCATT  
GTCACTGAGAGCAGGCTTTGGAGGCAGAGCCCCCAGTTGGAATCCCAACTTAACCA  
GCTAGGTTCCAGGTAGGACCCACAATTACCCGAGGAGAACAGTTGTGCCCTTCCCTGC  
AGGGCCAGTGTAAGAGTCCAGGAGTTAGTACACATAGAGATAGTGGCATGTGCTTTTTA

13086 GGCACGTGCCACACAACCCCTGGGTAAATTTTTTTTTTTTTTTTTTTGAGATAGGGTCTCTG  
TCTGTTGCCAGGCTGGTCTCAAATTCCTGGCCTCAAACCATCTCACACCTGAGGCGCT  
CAAAATATTGGGATTATAGGTGCGAGCCATCATGCTCAGCCAGAATAATAACTGGTTTTT  
TTTGTTTTTTTTTTGAGACAGAGTCTCACTCTATTACCCAGGCTCTGGAGGCCCAACTCG  
TGTTTGTTGTTATTTTATTTTATTTATTTATTTATTTTCGAGACAGAGCCTCTCTCTTT  
[T, C]  
ACCTAGGCTGGAGTGCAGTGGCGCAATCTCGGCTCACTGCAACCTCCGTCTCCTGGGTTT  
AAGTGATTGTCTGCCTCAGCCTCCTGAGTAGCTGGTGACAGGCGCGTGCCACCATGTC  
CCAGCTAATTTTTGTATTTTATGATAGACAGGGTTTTACTATGTTGGCCAGCTGGTTTC  
TAACTCCTGAACCTCGGGTATCTGCCTGCCTCGGCTCCCAAAGTGTGGGATTACAGGC  
ATGGGCTCCGTGCCCGGCCATGTATTTATTTAGGCAAGGTCTCTCTGTTATCCAGGC

13183 ACCATCTCACACCTGAGGCGCTCAAAATATTGGGATTATAGGTGCGAGCCATCATGCTC  
AGCCAGAATAATAACTGGTTTTTTTTTTGTTTTTTTTTTGAGACAGAGTCTCACTCTATTAC  
CCAGGCTCTGGAGGCCCAACTCGTGTGTTGTTGTTATTTTATTTATTTATTTATTTAT  
TTTCGAGACAGAGCCTCTCTTTCACTAGGCTGGAGTGCACTGGCGCAATCTCGGCTCA  
CTGCAACCTCCGTCTCCTGGGTTCAAGTGATTGTCTGCCTCAGCCTCCTGAGTAGCTGG

FIGURE 3, page 12 of 22

[T, C]  
GCTACAGGCGCGTGCCACCATGCCAGCTAATTTTGTATTTTGTAGTAGAGACAGGGTTT  
TACTATGTTGGCCAGCTGGTTTCTAACTCCTGAACTCGGGTGATCTGCCTGCCTCGGCCT  
CCCAAAGTGCTGGGATTACAGGCATGGGCCTCCGTGCCCGGCCATGTATTTATTTAGGCA  
AGGTCTCTCTCTGTTATCCAGGCTGAAGTGCAGTGGCACATTCATAGCTCACTGCAGCCT  
CAAATATCAAGTAACAGGGACTACAGGCATGCACCACACCCATCTACTTTTTTTT

21240 TCAGCTCACTGCAACCTCTGCCTCCCTGGCTTAAGCGATCCTCCCACCTCAGCCTCCTGA  
GTACGTGTGACCATAGGCCCATGGCACAAAGCCAGCTAATTTTTGTATTTTGTAGTAGA  
AATGTGGTTTCACCATGTTGCATAGGCTGGTCTCGAACTTCTGAACTCAAGTGATCTGCC  
TGCCTTGGCCTCCCAAAGTGCTGGGATTCTAGGTATGAGCCACCCTGCTCGGCCTATAAT  
GGCACTTTCTATCCCATGATGAGGCTCTACTCTCATGACCTAATCATCTCCCAAAGGC  
[C, G]  
CTAAGGCCTCTGTATACCATCACCTTTGGGGTTAGGTTTTAACATATACATTTTGGGGGG  
ACACAGACATTTTACACCATAGCACCTCCATTGAAAGGAAACATTTCTGACACCTGGCTA  
TCTCAAAGGGCCCTTTCACTTCCCCTGCAGGCTGCATTCCACATCACCAACAAGAGCAG  
CGACACTCACTCAGAGGTTAAATAACTTGTCCAGAGTCACAGCAGTAATGAATGACAGAG  
CTGGGGCTTGAATCCAGGCGTCTCCTAGAGCCTGGATTCTGTGTAGTGTAGTGAAGCTG

21695 CATTCCACATCACCAACAAGAGCAGCGACACTCACTCAGAGGTTAAATAACTTGTCCAG  
AGTCACAGCAGTAATGAATGACAGAGCTGGGGCTTGAATCCAGGCGTCCCTCTAGAGCCT  
GGATTCTGTGTAGTGAGTGAAAGCTGACTCCTGGGAGACTTCTGCGTGGTCTCTGGTTCTC  
TCTCCAGACTGCACCTGCCCAAGTTTCTTCTCCTGATGGTCCCTAGGCTATTACCAAGACA  
GTGGCCCTGCCTGTCAGGTGTTTTTATTACCAGATGAGGTCTAGGCTCAGGAACCTGT  
[A, G]  
GGAAGCTGAGTTCAGAGTCTTTGAGCAGGCTTTAGGGAGGTTCCAGCTTCCCACCACCAA  
GCCCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGTGGGGGGTCTGGAAGATGAGGTTG  
CGGGGTGCGATATTCTGCCCAATTTCGCCCTCCTTGCTCAATCTGTTTCTGCAGGTATTG  
CTGACTACAGACCCAAGGATGGAGAAACCATGAGCTGAGGCTGGTTAGCTGGTAGCCCC  
TGAGCTCCCTCATCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTGATG

22058 CCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGTGGGGGGTCTGGAAGATGAGGTTGCG  
GGGTGCGATATTCTGCCCAATTTCGCCCTCCTTGCTCAATCTGTTTCTGCAGGTATTGCT  
GACTACAGACCCAAGGATGGAGAAACCATGAGCTGAGGCTGGTTAGCTGGTAGCCCCCTG  
AGTCCCTCATCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTGATGTC  
CCTGGAACAGGAAGTCCGCTGACCCTGCTGCCACCTCCTGTGCACTTTGAGCAATGCCCT  
[C, T]  
TGGGATCACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTTGAGCAG  
AGAGCCCAAGCATCTTCCCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCA  
GGTCTCCCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGCC  
AAAAAACGGAGTCCGAGGCGCAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTC  
CTGCAAGAAGGCCCTCCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCT

22233 CCTGAGCTCCCTCATCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTG  
ATGTCCCTGGAACAGGAAGTCTGCCTGACCCTGCTGCCACCTCCTGTGCACCTTTGAGCAAT  
GCCCCCTGGGATCACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTG  
GAGCAGAGAGCCCAAGCATCTTCCCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGC  
CCTGCAGGTCTCCCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTC  
[C, A]  
TTGGCAAAAAACGGAGTCCGAGGCGCAGGTGTTGTGAAGACCACTCGTTCTGTGGTTG  
GGGTCTTGCAAGAAGGCCCTCCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACT  
CTGCTGTTAGAGTGGCAGCTCCGAGCTGGTGTGGCACAGTAGCTGGGAGACCTCAGCA  
GGGCTGCTCAGTGCCTGCCTCTGACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTAC  
AGTGGCCTGGTGCCTCATACTCCTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGGAAGTA

22245 TCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTGATGTCCCTGGAA  
CAGGAAGTCCGCTGACCTGCTGCCACCTCCTGTGCACCTTTGAGCAATGCCCTCGGGAT  
CACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGAGCAGAGAGCC  
AAGCATCTTCCCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCAGGTCTC  
CCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGCAAAAAA  
[C, -]  
GGAGTCCGAGGCGCAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTCTGCAAG  
AAGGCCTCCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCTGTAGAG  
TGGCAGCTCCGAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCAGGGCTGCTCAGT  
GCCTGCCTCTGACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTACAGTGGCCTGGTG  
CCTCATACTCCTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGGAAGTAACCCCATCAGGG

22375 ACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGAGCAGAGAGCCAAGCATCTTC  
CCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCAGGTCTCCCATGAAGGC  
CACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGCAAAAAACGGAGTCCGC  
AGGCCCGAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTCTGCAAGAAGGCCTCC

FIGURE 3, page 13 of 22

TCAGCCCGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCTGTTAGAGTGGCAGCTC  
[C, T]  
GAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCAGGGCTGCTCAGTGCCTGCCTCT  
GACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTACAGTGGCCTGGTGCCTCATACTC  
CTCAGGTGCAGGGGAGGACAAGAGAAGGGGGAAGTAACCCCATCAGGGAGGAGTGGAG  
GGTGCCTGAGCCGCATGTGGGCATTGGGGAGTGTATGGGAATGCCAGCAGTGTATGACGT  
TGACTACTGACTGAGCACCCTACTATGACTGAGCACTACTCGCTAGATACTATCTTG

23042 GCCGGGCGTGGTGGCTCACGCCTGTAATCCCAGCACTTTAGGAGGCCAAGGCAGGTGGAT  
CACAAAGTTCAGGAGTTTGAGATCAGCCTGGCCCAACATGGTGAACCTCCATCTTTACTAAA  
AATACAAAAAATTAGCCAGGCATGGTGTGTCATGCCTGCATGCCTGTAATCCCAGTTACT  
TGGGAAGCTGAGGCAGGAGATTGCTTGAACCTGGAGGCGGAGGTTGTAGTGAGCCGAG  
ATCAGCCCATTCAGCTCCAGCTTGGGCAAGAAGAGAACACTCTCAAAAAAAAAAAAAA  
[A, -, T]  
CAGGAAACTGGTGTCAAAAAGGAAAAGTGAAGTCAACCAAGGTCACAGACTAGGCAGTGAT  
GCTGGGGGAACCTGGCTCAGGGGACACAGACCTGGCCTGGGGCAGCCTTGACAGCTCCTCC  
ACTAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAA  
CTCAACTGGAGCCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTTCCACAAGGCA  
CCATTAGGGCATCACCCAGGCCCTCGGCAGCCACGACGCAGGGATCCTGCCTCTCATTTGG

23344 AGGAAACTGGTGTCAAAAAGGAAAAGTGAAGTCAACCAAGGTCACAGACTAGGCAGTGATG  
CTGGGGGAACCTGGCTCAGGGGACACAGACCTGGCCTGGGGCAGCCTTGACAGCTCCTCCA  
CTAAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAAC  
TCAACTGGAGCCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTTCCACAAGGCAC  
CATTAGGGCATCACCCAGGCCCTCGGCAGCCACGACGCAGGGATCCTGCCTCTCATTTGGT  
[T, C]  
GGGGGCTTAGGGGCTCTGGGCTGCCCTCTTGAAGAGGGGTTTCAGCCAGCGAGGCACCC  
CCTATGCTGCACCCACCAAGGTTAGGAAGAGGTCCTGTCTCAGTGGGGCCCTCTGATG  
AACAGCCCATCAGGTCTGCGTCCACATGCCTTGAAGAGATGGTGACATACTCAAAGTCC  
TTGAAGCCGCATATTAAACACCTAGAGCACCATTCTCAAACATTTAGGGTCTGAGAAGA  
TAGGGGAAGTAAGCAATTTAAACATTTCTTTATATTGGGCCAGGTGCAATGGCTCACGT

23873 GGTCTGAGAAGATAGGGGAAGTAAGCAATTTAAACATTTCTTTATATTGGGCCAGGTGC  
AATGGCTCACGTCTGTAATCCCAGCGCTTTGGGAGGACGAGGATCAGCTGAGGTGAGGAG  
TTCAAAATCTGAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAAC  
CTCAGGCGTGGTGTATGTACCTGTAATCCTAGCTATTTCAGGAGGCTGAGGCACAGAAT  
TGCTTGAGTCAATATTGACCACTGCACTCCAGCCTGGGCAACAGCGAGACTCTTGTCTC  
[A, -]  
AAAAAAAAAAAAAGATATTGCTGAAAAGACCCAGCCTGCCAAACTCAGGGGAGCCAAGG  
GAGGTAGTGAAATGGAAGTTGGAGCTCAGCGCTCCACACCTCCACTGCCCTCAGGCCTT  
CTCTGCCCTCTTTCCCATCAGTCAGCTGCTTCTGGGCATGGTCTGGCAGAGACTTGGCCT  
CCTTCCAGTTCAAGCTCCCTCTTAGATTGTGTCCACGCCACTGAGTCTTTGGGACACTG  
GGTCAGATGTCTAGTCTGGCACAATTGGCAGGAATCCCAAGAAACAGTGTGAGTGAGGGG

24764 ATAGCTGGGATTACAGGTGTGCACCACCATGCCAGCCTAATTTTGTATTATTAGTAGA  
GATGGGGTTTACCATTGTGTCCAGGCTGGTCATGAACCTCCTGACCTCAAGTGATCCACC  
CGCTTTGGCCTCCCAAGTGTGGGATTACAAGCATGAGCCACAGTGCCTGGCCTGACCC  
TGCTCTTTTGAAGACCATTCCCCAAATTCGTGCACCTGTGTGCCTTTCTTCTCTCTG  
CCTCCTCTCAGCTCTGCCCGCTCTCCTCCCTTCTCCTCTGGCAATCCCACTCATCTCT  
[G, T]  
GAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGGATGAAG  
GACGTGGCCACCGAGATTGTTTGTGTTTGTGTTTGGAGATGGAGTTTGTCTCATGTGCCC  
AGGCTGGGGTACAATGTTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGTTCAAGC  
GGTTCTCCTGCCTTAGCCTCCCAAGTAGCTGGGATTACTGGCATGAACCACCACCTGG  
CTAATTTTGTGTTTTTAGTAGAGATGGGGTTCTTCATGTTGGTCAGGCTGGTCTCGAAC

24939 GACCCTGCTCTTTTGAAGACCATTCCCCAAATTCGTGCACCTGTGTGCCTTTCTTCT  
CTCTGCCCTCCTCAGCTCTGCCCGCTCTCCTCCCTTCTCCTCTGGCAATCCCACTCA  
TCTCTTGAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGG  
ATGAAGGACGTGGCCACGGAGTTTGTGTTTGTGTTTGGAGATGGAGTTTGTCTCATG  
TTGCCAGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGT  
[T, C]  
CAAGCGGTTCTCCTGCCTTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCACCACA  
CCTGGCTAATTTTGTGTTTTTAGTAGAGATGGGGTTTCTCATGTTGGTCAGGCTGGTCT  
CGAACTCCCAACCTCAGGTGATCTGCCTGCCTCGGCCTCCCAAGTACTGGGATTACAGG  
GTGAGGCACTGTGCCTGGCCAGGCCCACGGAGTTTAAAGAGGCTTCCTGTGGCAGTGG  
CATCCAGACGGAGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGGAGTG

24945 GCTCTTTTGAAGACCATTCCCCAAATTCGTGCACCTGTGTGCCTTTCTTCTCTCTGC  
CTCCTCTCAGCTCTGCCCGCTCTCCTCCCTTCTCCTCTGGCAAATCCCACTCATCTCTT  
GAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGGATGAAG

FIGURE 3, page 14 of 22

GACGTGGCCACGGAGTTTGTGTTTGTGTTTGTGAGATGGAGTTTGTCTCATGTTGCC  
AGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGTTCAAGC  
[G, A]  
GTTCTCCTGCCTTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCACCACACCTGGC  
TAATTTTGTGTTTGTAGTAGAGATGGGGTTTCTTCATGTTGGTCAGGCTGGTCTCGAACT  
CCCAACCTCAGGTGATCTGCCTGCCTCGCCTCCCAAAGTACTGGGATTACAGGGTTGAG  
CCACTGTGCCTGGCCAGGCCACGGAGTTTAAAGAGGCTTCTGTGGCAGTGGCATCCA  
GACGGAGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGAAGGGGAGTGTGAGTT

25092 ATCATGCTGCTTCTCCTGTGGGAGGGATGAAGGACGTGGCCACGGAGTTTGTGTTTGT  
TTGTTTGTAGATGGAGTTTGTCTCATGTTGCCAGGCTGGGGTACAATGGTACGATCTCA  
GCTCACTGCAACCTCTACGTCCCGGGTTCAAGCGGTTCTCCTGCCTTAGCCTCCCCAGTA  
GCTGGGATTACTGGCATGAACCACCACCTGGCTAATTTTGTGTTTGTAGTAGAGATGG  
GGTTTCTTCATGTTGGTCAAGGCTGGTCTCGAACTCCAACCTCAGGTGATCTGCCTGCCT  
[C, T]  
GGCCTCCCAAAGTACTGGGATTACAGGGTTGAGCCACTGTGCCTGGCCAGGCCACGGGA  
GTTTTAAAGAGGCTTCTGTGGCAGTGGCATCCAGACGGAGTGCAGAACTCAAAGTTGAA  
GGCCAGAAGCTCAGGGAAGGGGAGTGTGAGTTGAGGAGTCTCTTGGCTGCCAGGGCCAG  
AAACCGAACTCCAAGCTCTCCACAACAGCGGGTGTAGAGCATGTAGAATCAGAGAGGAG  
GCTGAGCCATGCAGCCCCGAGAAGAGGGGAATGCCACTGAGCCACAGAGACCCAGTGCCA

25428 AGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGAAGGGGAGTGTGAGTTGAGGA  
GTCTCTTGCTGGCTGGAGGCCAGAAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAG  
AGCATGTAGAATCAGAGAGGAGGCTGAGCCATGCAGCCCCGAGAAGAGGGGAATGCCACT  
GAGCCACAGAGACCCAGTCCACTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCC  
[T, G]  
GCCTGTCTCTGTATGCAGGCTTCAACCTCTCTCGTTGTACATTGTACACATTCTAGGTGA  
CACCAGCAGCTTCTGATTCTCATCTCCATAACATCAGCCCCCAGAGAGGGGACAACCTG  
CTGAGCTGATAACATAATAGATGCCCTTTCTTGAGGCCATGTCATGGTCAGCGTGGA  
GAGGATGAAGCCTGAGCAGGCAGGATCGGGGCTTAGAGGGGAAGGAGGTGGAAGTT

25513 GGCCAGAAGCTCAGGGAAGGGGAGTGTGAGTTGAGGAGTCTCTTGGCTGCCAGGGCCAG  
AAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAGAGCATGTAGAATCAGAGAGGAG  
GCTGAGCCATGCAGCCCCGAGAAGAGGGGAATGCCACTGAGCCACAGAGACCCAGTGCCA  
CTGCCAGGTGTCTCTGCCTTCACTTCCATGACCCGGCCTGTCTCTGTATGCAGGCTTCA  
CCCTCTCTCGTTGTACATTGTACACATTCTAGGTGACACCAGCAGCTTCTGATTCTCATC  
[C, T]  
CCCATAACATCAGCCCCCAGAGAGGGGACAACCTGTGAGCTGATAACATAATAGATGCC  
CCTTTCCTGGAGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGAGCAGGCAGGA  
TCGGGGGTCTAGAGGGGAAGGAGGTGGAAGTTGAGATCACAGACCTGTGGTCAGGTGGCC  
TGGAAGGGTTTGACGAGTGTGCGCCCAAAGAGCTTGGAAGGGATTTTGTCTGTGTGGGT  
GAGCACTGCCTCTCCCTTAGGGACAACAGCCACCTCTTCTCTCCCATTTGCCCTTTCCC

25684 CCAGTGCCACTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCCGGCCTGTCTCTGTATG  
CAGGCTTCAACCTCTCTCGTTGTACATTGTACACATTCTAGGTGACACCAGCAGCTTCTG  
ATTCTCATCTCCATAACATCAGCCCCCAGAGAGGGGACAACCTGTGAGCTGATAACAT  
AATAGATGCCCTTTCTTGAGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGA  
GCAGGCAGGATCGGGGCTTAGAGGGGAAGGAGGTGGAAGTTGAGATCACAGACCTGTGG  
[C, T]  
CAGGTGGCCTGGGAAGGGTTTGACGAGTGTGCGCCCAAAGAGCTTGGAAGGGATTTTGTCT  
GCTGTGGGTGAGCACTGCCTCTCCCTTAGGGACAACAGCCACCTCTTCTCTCCCATTT  
GCCTTTCCCTTCTGTAGATAGAAACACAGGCCTCTTGTGAGGCCCTTCTTAACTTCC  
GTGATGGGGAAGCGGCCGAGAAAGGGAGTTCTGGCAGCTTCTCCGAGACCCCAACACC  
CCACTGTTGCAAGGTGAGTATGGCCTGACACTCTGGATGTGTCCCTACCCCAAGCTTA

26165 GTGATGGGGAAGCGGCCGAGAAAGGGAGTTCTGGCAGCTTCTCCGAGACCCCAACACC  
CCACTGTTGCAAGGTGAGTATGGCCTGACACTCTGGATGTGTCCCTACCCCAAGCTTA  
CTCAGCCAAGAGGCTTCATCAACTCAGCCAGCTTCCCTAGCACCTCTCTGGGCCACAC  
CTTCACAAAATCACTGATGCTCAAAGTTGGATATAATATATTGAACTGAAGCCTTAGCAT  
TTTTATGCAAGTTACTGTGGAATTTCTAGGAAACCAGACAGATTACAAAAA  
[A, -]  
CTAGAAGAAAATTAACATCACCTAGGATATACTACCTAGGAATAACGTCTTTTATTTTGA  
GATGGAGTTTTCGCTCTTGTGCCCCAGGCTGGAGTGCAGCGGTATGATCTCGGCTCGCTGC  
AACCTCCGCCTCTCGGTTTCATGTGATTCTTCCACCTCGGCCTTCTAGAGCCCAAGTGG  
TCTGCCTGCCTCTGCCTCCCAAAGTTCTGGGATTACAGGCATGAGCCACCGCAGCCAGCC  
AAAATTACTTAACCTTCTCTAGATACTTTTAAAAATATGGCAGTAAGTTTTTCATAA

Isoform 2:

FEATURES:

Exon: 2132-2195  
Intron: 2196-5670  
Exon: 5671-5863  
Intron: 5864-7672  
Exon: 7673-7761  
Intron: 7762-9149  
Exon: 9150-9302  
Intron: 9303-10101  
Exon: 10102-10274  
Intron: 10275-10399  
Exon: 10400-10586  
Intron: 10587-12128  
Exon: 12129-12294  
Intron: 12295-25922  
Exon: 25923-26040

Allelic Variants (SNPs):

DNA				Protein		
Position	Major	Minor	Domain	Position	Major	Minor
1022	C	T	Beyond ORF(5')			
1882	C	T	Beyond ORF(5')			
1951	G	A	Beyond ORF(5')			
2940	A	G	Intron			
3831	G	A	Intron			
6732	G	A	Intron			
7558	G	A	Intron			
7931	T	A	Intron			
8190	T	C	Intron			
8652	C	T	Intron			
9370	G	C	Intron			
9463	C	T	Intron			
9883	G	T	Intron			
11594	G	A T	Intron			
12361	A	G	Intron			
13187	T	C	Intron			
13284	T	C	Intron			
21341	C	G	Intron			
21796	A	G	Intron			
22159	C	T	Intron			
22334	C	A	Intron			
22346	C	-	Intron			
22476	C	T	Intron			
23143	A	- T	Intron			
23445	T	C	Intron			
23974	A	-	Intron			
24865	G	T	Intron			
25040	T	C	Intron			
25046	G	A	Intron			
25193	C	T	Intron			
25529	T	G	Intron			
25614	C	T	Intron			
25785	C	T	Intron			
26266	A	-	Beyond ORF(3')			

Context:

DNA  
Position

1022 TTGGAGATATTTTAAGGTCATAGTGTCTTCACAAATTGAGCTGAAAGGGAAGTGTAGGA  
TGATCTTGCCTAACCCCTCTCATCTCACACAGGAAGAACTATTTTAAACTCGAGAGGTTAA  
GTGACCTGGCCAAAGTCACACAGCCACCACTAGTTAACTCGTATACATTGATTCTCCTGT  
GGGGCTGGGCAGATGAGGAATCTTTTGTCTCTTCCCTGTTTGCAGAGATTTTTTTGAG  
GTTACTTTCCGAGTTCTGGCAAGTACCCCTGCTTCTGGTAGCTTTGTGTCTCGATTCAAT  
[C,T]  
TCATTCTTTTATTTTATTTTATTTTATTTTGGAGACAGGGTCTCACTTTGTCAACCAAGCTGGA



GTGCAGTGGTGTAACTCTTGGCTCACTGTAGCCTCCACCTCTTGGGTTCAAGCGATCCTCC  
TGCCTCAGCCCCCAAGTAGCTGGGATTACAGACGTCTGCCACCACGCCAGGCTAATTTA  
TGGTTTTTTGTATGTGTTTTTTGTGTTTTTGTAGAGACAGTGTTCCTCATGTTGCCAG  
GCTGGTCTCCAACCTCTGAGCTCAAGTGATCTGCCCGCTCAGCCTTTCAAAGTGCTAGG

1882 ACAGCTGACTCCAGCAATGTGCTCACGTGACCACTGCAGCTGCAGCTCCCGTTCCACTC  
CTTGTCTCTGGGCTAGGTGGGCACTACCAGGGGCTCCTTTGGTAAGGAGTACCGGGTAGGC  
ACCCGGTCTCTGCCAATCCACCACTGGAACAGCTGGGGGGACAGCAGACAGGCACGGTCGG  
ACAGACTTGACAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCGGGCTCTTTAAGCAGGA  
ACGTGAATGGCCTCAAGATGTCTCACATGGTCCCCTAGCCCTCCTCCTCCTTTGTTCC  
[C, T]  
TACCTCCAGGAGGGCTGCTCTGCCCTTCCTTCTCTGTTCTTTGGCCTTATGTTCCCCGC  
CACCACAGGCCTTCCCCCGCCCCACCCCTCTGCAGACTTAGCCGTGCATTGCAGGCATGG  
AGGATTAATCAGTGACAGGAAGCTGCGTCTCTCGGAGCGGTGACCAGCTGTGGTCAGGAG  
AGCCTCAGCAGGGCCAGCCCCAGGAGTCTTTCCGATTCTTGCTCACTGCTCACCACCT  
GCTGCTGCCATGAGGCACCTTGGGGCTTCCTCTTCTCTGTTCTGTTGGGGTCTGGGGCCCTC

1951 GGCTAGGTGGGCACTACCAGGGGCTCCTTTGGTAAGGAGTACCGGGTAGGCACCCGGTCC  
TGCCAATCCACCACTGGAACAGCTGGGGGGACAGCAGACAGGCACGGTCGGACAGACTTG  
ACAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCGGGCTCTTTAAGCAGGAACGTGAATG  
GCCTCAAGATGTCTCACATGGTCCCCTAGCCCTCCTCCTCCTTTGTTCCCTACCTCCA  
GGAGGGCTGCTCTGCCCTTCCTTCTCTGTTCTTTGGCCTTATGTTCCCCGCCACCACAG  
[G, A]  
CCTTCCCCCGCCCCACCCCTCTGCAGACTTAGCCGTGCATTGCAGGCATGGAGGATTAAT  
CAGTGACAGGAAGCTGCGTCTCTCGGAGCGGTGACCAGCTGTGGTCAGGAGAGCCTCAGC  
AGGGCCAGCCCCAGGAGTCTTTCCCGATTCTTGCTCACTGCTCACCACCTGCTGCTGCC  
ATGAGGCACCTTGGGGCTTCCTCTTCTTCTGTTGGGGTCTGGGGCCCTCACTGAGATG  
TGTGGTGAGTAACCTCGCCTCTATCTGTGCTCTTTCTCTGGGTCCTTAGTGGGGTGG

2940 AACATAGGGAGACCCCATCTCTACAAAAAATAAAAAAATAAAAAATAGCTGGGCATGG  
GGAAGACTTTCTGAAGACCAAGAGGACACATGGGAGCTGAAACTCGAAGGAAGAAAAGGA  
GCTGGCAGGAAGGAGTGGGGGACACACATTCTAGGCAGCAGGAAGTGAACCTTCGGAGG  
TCCTGCTGCTCCAGCTCTGTGCCCAAGGGGTCTTTGGAGCACAGTCTCCTGGGACCT  
GTCTATGAGTCTGAGCTTAGAGGCTCAGGGCTGCTCCTTCAGACAGGAGGCAGAAGGCAG  
[A, G]  
CTTTGGGAACCTTTGGGCGGCCACGCGCCTTTTCTCCTCCTCTGCACCTAGGATTACGTT  
GAGCAATACACTTTACCCCATGCTCTCTTGAAGCCTGGGGAAACCTGAGAGGTGGG  
TGCAGTCACTGTCAGGTGTCAAGTGAAGAAGTCAAGGGTGGAGGGGCTGAGTGACCCAC  
TCAGGTGCTTCCACTTTTCCAGAGCTTTGCTGAACCTAGTTTTTAGAAGCTGAAGCCTC  
GTTTGTTCGTTTTGTTTTTGTGTTGAGAGAGGTCTCCCTCTGTTGCCAGGCTGGAGT

3831 GACACCTCAGGTCTGGGCCCAGGAACCCAGCTCTTGGTTCATGTCCGGACAGTCCCCAG  
GGGAGTTCTGGGTTCAACCAGCAAGAGCTCTTCTCCTGGCTGATCTGGTCTCAGCCTT  
GGACAGTTAGTCCATTAACCTGACCCACAGGAGCCCCAATCCTTTGGGGTCTGGGGAAT  
CTTGAAGTGGGTTTGGGGTGCAAAATATCTGCACTGAGTCACTTAATTGCACCCAGCCTC  
ATTCTTTTATCTGTAAAGTGGGCTAAGAATGCTCCCTGCCTTCTCCTCGGTGTAGTAC  
[G, A]  
AGGAAGGATCCCATGACACCTGCTCTCCAGTTTAAAGCTCTATATGTATGTTGTGAAAT  
TGACAGGGATCGCTGCACAAACGCTAATGCAAAGTGGGCTCCTGTGCTTCTTTCTCTT  
TCTTCTTCTTTTTTTTTTTTAAATTTTCTTCTAGAGATGAGGTCTCACTATATTGCCCA  
GGGTTGGTTTCAAACTCCTAGGGTCAAGCGATCCTCCACCTTGGCCTCCCAAACTGCTG  
GTATTACAGGCGTGAGCCACTCTGTCTGGCTCCTATGCTTGTGAATGTCAACAGCAATCA

6732 TGAATAGCTGGGATTACAGGCGTGTGCCACCATGCCAGCTGATTTTTGTATTTTTTGGTA  
GAGATAGGGTTTACCATGTTGGCCAGGCTGGTCTTGAACCTCCTGACCTCAGGTGATCCG  
CCTGTCTTGGCTTCCCAAAGTGTGGGATTATAGGCATGAGCCACTGCACCAATCCAAAA  
GCAGCATCTTTGTGCTCCCTTTTCAAGAGGCATCAGAGAGGCGCTGTTTGGGGTTTGA  
ATGAGAGGGCGAAGATCAGCCATGGAGTGCCCTTTCTCAGACTCCCTCTTGAGAAGTGG  
[G, A]  
TGCAGGGGTGGAGAGAAAAGAAGACTAGGCATAGTGGCTCATACCTGTAATCCCAACATT  
TTGGGAGGCTGAGGCAGGAAGATTGCTTGAAGCTCAGGAGTTTGAAGCAGCCTAGGCAAC  
ATAGTGAGACCACATCTCTTAAAAAAGAAAAAAGAAAAAATGAGCCAGGTGTAGTGA  
CTCATGCTGTGGTCCCCACTTCTCCGAGGCAAGGTGGGAGGATCTTTTGAAGCTGAG  
AAATCAGAGGTACAGTGAGCCATGGTGGCACCCTGCACTCCAGCCTGGGAGACAGAGAG

7558 AAGAATCGCTTGAACCTCGGGAGGCGGAGGTGTCAGTGAGCTGAGAACATGCCACTGCACCT  
CCAGCCTGGGCAACAAGAGCGAAACTCTGTCTCAAAGAAAATAAATAAATAAATAAATAA  
AATAAAAAAGGAGGGGGCATATGGGTGAAGTATGGACAAAATAGTGGGGCAGGCACAGAT  
GATCTGGACACAGGAGCCCTTGGAGTTTATTCTTGAATCTAACTGTTTCATCTTTATTA  
TATTTGTGGCATACACCTCACAACAACATAGCCAACACACCTCCTTTTGGAGCTTTATC  
[G, A]

FIGURE 3, page 17 of 22

AAGTTTCCCAGTGTTAAGATTTTTTCCCCTTTGTGATGCGGGTGGGGTGGGTGCTGTAA  
 GCAGGCTTACGGGGTGGCAGTTTCTCACAAGGCATTAACTGCGCTTGTCTAGGTCTGC  
 CTTACAGCGAGGATGACGGTGACTGCCAGGGCAAGCCTTCCATGGGCCAGCTGGCCCTCTA  
 CCTGCTCGCTCTCAGAGCCAAGTGTGAGTTTGTAGGGGCCACAAGGGGGACAGGCTGGT  
 CTCACAGCTCAAATGGTTCTCGAGGATGAGAAGAGAGCCATTGGTGAGCAGACACCATC

7931 GGTGGCAGTTTCTCACAAGGCATTAACTGCGCTTGTCTAGGTCTGCGCTTACGCGAGGA  
 TGACGGTGACTGCCAGGGCAAGCCTTCCATGGGCCAGCTGGCCCTCTACCTGCTCGCTCT  
 CAGAGCCAAGTGTGAGTTTGTAGGGGCCACAAGGGGGACAGGCTGGTCTCAGAGCTCAA  
 ATGGTTCTTGAGGATGAGAAGAGAGCCATTGGTGAGCAGACACCATCCGCTGGGGGTGG  
 GGAGCAGCTGGGAGGGCTCATCAGATGATATTCTCCAATGAGAATCAGAAGTTGGGTTT  
 [T, A]  
 CTCCTCCAGGCGTCTTCCCACCATCCATTCTGCCCATCTCACTGCGCTACGTAGAGGCTCG  
 AACCTGTCCCAGTATCCATCCTTGACCCAGCTTTTCCCGCGCTGCACACATACTATTGAC  
 AGGTGTGTTTCGTGGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG  
 TCTTGTCTGCCAGGCTGGAGTACAATGGCGCAATCTCAGCTCACCGCAATCTCTGCTCC  
 TGGGTTCAAGCAATCTCTTGCCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCATGCGCC

8190 ATCAGATGATATTCTCCAATGAGAATCAGAAGTTGGGTTTCTCCCAGGCGTCTTTCC  
 CACCATCCATTCTGCCCATCTCACTGCGCTACGTAGAGGCTCGAACCTGTCCCAGTATGCA  
 TCCTTGACCCAGCTTTTCCCGCGCTGCACACATACTATTGACAGGTGTGTTTCGTGGTTT  
 TTTGTTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT  
 AGTACAATGGCGCAATCTCAGCTCACCGCAATCTCTGCTCCTGGGTTCAAGCAATCTC  
 [T, C]  
 TGCCTCAGCTCCTGAGTAGCTGGGATTACAGGCATGCGCCACCACACCCAGCTAATTTT  
 GTATTTTGTAGTAGAGTGGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAAGTCTTGACC  
 TCAGGTGATCCGCTTGCCTTAGCCTCCGAAAGTGTGGGATTACAGGCATGAGCCACTGC  
 GTTAGGCCCACTGACAAGCCTTGTATTGGCTAGCCACCAAGATTGACTTGATTATCCACC  
 TTCGGGACAAGTGGACAGCTGCTTATGACTTACGCCATAGTCTGTCTCTACTAGCTCTC

8652 TACAGGCATGAGCCACTGCGTTAGGCCCACTGACAAGCCTTGTATTGGCTAGCCACCAAG  
 ATTGACTTGATTATCCACCTTCGGGACAAGTGGACAGCCTGCTTATGACTTACGCCATAG  
 TCTGTCTCTACTAGTCTCTCTGCGCTGACTTGACCCAGCATAACAAGCCAGAGCCAGCC  
 TTTTCAATATAAAGCTGATCTTGTGCTGGCACTGCTTAAACCTGCGAGGGGCTCGCACTGC  
 TCCATGGCCAGCCTGTCTACCTTACCTTCTGCCCAGGCTCTGCTCATCCATTCTCTGC  
 [C, T]  
 TCCCACACACTGCGCTCTGTGGGCTCCAGCCATACCATCTCTCAACTCATAAGCCAGTT  
 TTTTCATACAGGCTCCCTCCATCTGAGTGGCTTCCCTGCGTGCAGTTCACTCCTGCTCT  
 ACCTTGGCTCTGCTCCACCCATCCTCAGCCGTCTCCAGCATTACCTCCTTGAGAGTCT  
 CTGCTTGTACTTCCAGCCACCCAAATATCACTACTTGGTCTGCATTCTCGTTGCAATTG  
 CAGTCGATGAGCAATTGCTGTGGTTGAGGCCCCGAAGTGGCAAGTGCCTGTCTGCCATG

9370 AGGCCAGGGTCCCAGGTGCTGGCGGGGCTGGCTGCTGGGTGGGGGCAGAGAGGCAACCCC  
 TCTGTTTTTTTCCCTCTCAGGGCATGATCACAAGGGGCCACCCACACTAGCTACTACCA  
 GTATGGCCTGGGCATTCTGCGCTGTGTCTCCACCAGAAGCGGGTCCATGACAGCGTGGT  
 GGACAAGTCTGTATGTGTGGAACCTTCCACCAGGGGCCACCATCTGTGGGTGAGTA  
 GGTACAGCGTGCCAAGGCCAGGCTGGCACTCCCTCAGTCCCAGGCTGCACTGATGAC  
 [G, C]  
 TCCATACCTTGGCCCCACACTCACCTTTCCTTGGGGCTCCTCCGAATCAAGTCCCTTAG  
 GGACGAATTGGCGAGGGCTCATGGGTGATGCTCCAGCTGTGAGCCAGCTTTGGAGCTGGT  
 AGGTGGATCTCTTGAGGCCAGGAGTTCAAGACAACGTGGTGAACCCCATCTCTACTAAA  
 AATAAAAAAGTTAGCCGGGCATGGTGGCACATGCCTGTAGTCCAGCTACTCGGGAGGCT  
 GAGGCAGGAGAATCACTTGAACCTGGGAGGCGGAGGCTGCAGTGAGTGAGATCGCACCA

9463 GGGCCACCCACACTAGCTACTACAGTATGGCCTGGGCATTCTGGCCCTGTGTCTCCA  
 CCAGAAGCGGGTCCATGACAGCGTGGTGGACAAACTTCTGTATGCTGTGGAACCTTTCCA  
 CCAGGGCCACCATCTGTGGGTGAGTAGGTGACACCGTGCCAAGGCCAGGCTGGCACTCC  
 CTCAGTCCCAGGCTGCACTGATGACGTCCATACCCTGGCCCCACACTCACCTTTCCT  
 TGGGGCTCCTCCGAATCAAGTCTTTAGGGACGAATTGGCGAGGCTCATGGGTGATGCT  
 [C, T]  
 CAGCTGTGAGCCAGCTTTGGAGCTGGTAGGTGGATCTCTTGAGGCCAGGAGTTCAAGACA  
 ACGTGGTGAACCCCATCTCTACTAAAAATAAAAAAGTTAGCCGGGCATGGTGGCACATG  
 CCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCGG  
 AGGTGCACTGAGTGGAGATCGCACCACTGCCCTCCAGCCTGGGCAACAGAGTGAGTGAG  
 ACTCTGTCTCAAAAAATAAAAAATAAAATAAACTCCCTAGTGATTCCAATGTGCAGCT

9883 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCG  
 GAGGCTGCAGTGAGTGGAGATCGCACCACTGCCCTCCAGCCTGGGCAACAGAGTGAGTGA  
 GACTCTGTCTCAAAAAATAAAAAATAAAATAAACTCCCTAGTGATTCCAATGTGCAGC  
 TAAGTTTGGAATAGGTGGTATGGGGTCAAGTCCTCTTGGGCTCCCTCCTCCAGTCTCT  
 CTCCTAACCTCTAGCCCTCAAGTTGCAGAGTGATCAGCCAACAGGTTTGGCCAGAAAT

FIGURE 3, page 18 of 22

[G, T]  
AGCAGTTTCCTGGGACACAGGATTTTCAGAGTCCAGACAAGGAAAGTCTTGGGCAGACCA  
GGTTGAGTTGGTGCCCTTAGCTGATCTGACCATGTTGCCCTTCTTCTCCAAGCCCTCTG  
TGGTTGTCCATAGCTACAAGGCCCTGACCCCTCAAGCCCTGCCTGTCTGGCCCTTTGG  
CTCTCCAGCTCATTGCATGTTCTGTCCCCCACTTCAAGACACAGCAGCCATGGCAGGCTT  
GGCATTACCTGTCTGAAGCGCTCAAACCTTCAACCTGGTTCGGAGACAACGGATCACCAT

11594  
AAAAAAGAGTGGAGAAGAAGGAGCTGGACATGGTGGCTCGTGCTTATAATCCTAGCA  
CTCTGGGAAGCTGAGGCAGATGGATTGCCCTGAGCCAGGAGTTTGGAGCCAGCTGGGCA  
ACATGGTGAAACCTGTCTTTACTAAAATACGAAAGATTAGCCAGGCATGGTGGTAGACA  
CCTATAATCCAGCTACTAGGGAGGCTGAGCCACAAGAATCACTTGAACCTGGGAGACAG  
AGGTTGCAGTGGCCGAGATCGCGCCATTGCACTCCAGCCTGGCGACAGTGTGAGACTC  
[G, A, T]  
GTCTCCAGAAAAACAAGAATGGATAGAGTGGAGCCAAGAAGAGGCAGGAAGAACAAGA  
CACAGAGGTGCACAGAGTTTGGGGGAATTTGAGGAATGGTCTTGCAAAAGAGTGGGATC  
TGGGAGAATGAGTGGGAGTGGAAAGCAGATGAATGAAGAGAAGGTGAGCGCATCAGGGTA  
ACAGAGATGCGTTGTGAACAAATGCATGTTCTAGGAAGAGCCCTCTGGAGTGTAGGTGC  
CAGAGAGGTGGGAGGAAGGATACTGGAAGCAGAGAAACAGTGGGGGCTGATCTTGGG

12361  
ACAAGAAGACAAATAATCCAGGCTCTCTGTCTCACACCAGCTGCCCCCCCCCTTTCTTCC  
TGGCACAGTCAATGTTGGAACAGCTGCTGAGACCATTCCTCAGACCCAGAGATCATCAG  
TGTACAGCTGCAGGTGCTTAGTCTCTTGGCGCCGTACAGACAGTCCATCTCTGTTCTGGC  
CGGGTCCACCTGGAGAGATGCTCTGAAGAAGGCCATGAGTTAGGAGGATTCAGGTGAGA  
CTCCACCTCCAGTCTCACCCCAACCTCACATGCCTGATAACAGGGTCACAGAA  
[A, G]  
AGACGGGGAACAGAGGAGAGGGTTCCCTCGGGAGAGACACTGGCCCTGCTTCTGCTTCTA  
CCTGCTCAGCTCCTTTCTTGGCCACGGTGTATGGAAACAGGGAGCCATAGGCCAGCATT  
GTCACAGAGAGCAGGCTTTGGAGGCAGAGCCCCCAGTTGGAATCCCACTCTAACCA  
GCTAGGTTCCAGGTAGGCACCCACAATTCACCGAGGAGAACAGTTGTGCCCTTCCCTGC  
AGGGCCAGTGTGAAGAGTCCAGGAGTTAGTACACATAGAGATAGTGGCATGTGCTTTTAA

13187  
GGCACGTGCCACACAACCCCTGGGTAATTTTTTTTTTTTTTTTTTTTGGATAGGGTCTCTG  
TCTGTTGCCAGGCTGGTCTCAAATTCCTGGCCCTCAAACCATCCTCACACCTGAGGCGCT  
CAAAATATTGGGATTATAGGTGCGAGCCATCATGCTCAGCCAGAATAATACTGGTTTTT  
TTTGTTTTTTTTTTTGAGACAGAGTCTCACTCTATTACCAGGCTCTGGAGGCCAACTCG  
TGTTTTGTGATTTGTTTTATTTTATTTATTTATTTATTTTCGAGACAGAGCCTCTCTCTT  
[T, C]  
ACCTAGGCTGGAGTGCAGTGGCGCAATCTCGGCTCACTGCAACCTCCGTCTCTGGGTTT  
AAGTGATTGTCTGCTCAGCCTCCTGAGTAGCTGGTGCTACAGGCGCGTGCCACCATGC  
CCAGCTAATTTTTGTATTTTATAGTAGAGACAGGTTTTACTATGTTGGCCAGCTGGTTT  
TAACTCCTGAACTCGGGTGATCTGCCTGCCTCGGCCCTCCAAAGTGTGGGATTACAGGC  
ATGGGCTCCGTGCCCGGCCATGTATTTATTTAGGCAAGGTCTCTCTCTGTTATCCAGGC

13284  
ACCATCTCACACCTGAGGCGCTCAAAATATTGGGATTATAGGTGCGAGCCATCATGCTC  
AGCCAGAATAATACTGGTTTTTTTTTTTGTTTTTTTTTTGAGACAGAGTCTCACTCTATTAC  
CCAGGCTCTGGAGGCCAACTCGTGTGTTGTGATTTGTTTATTTTATTTATTTATTTAT  
TTCGAGACAGAGCCTCTCTCTTTACCTAGGCTGGAGTGCAGTGGCGCAATCTCGGCTCA  
CTGCAACCTCCGTCTCCTGGGTTCAGTGATTGTCTGCCTCAGCCTCCTGAGTAGCTGG  
[T, C]  
GCTACAGGCGCGTGCCACCATGCCAGCTAATTTTTGTATTTTATAGTAGAGACAGGGTTT  
TACTATGTTGGCCAGCTGGTTTCTAAGTCTGAACTCGGGTGATCTGCCTGCCTCGGCCT  
CCCAAGTGTGGGATTACAGGCATGGGCTCCGTGCCCGGCCATGTATTTATTTAGGCA  
AGGTCTCTCTGTTATCCAGGCTGAAGTGCAGTGGCACATTATAGCTCACTGCAGCCT  
CAAATTATCCAAGTAACAGGGACTACAGGCATGCACCACCACCCATCTACTTTTTTTTT

21341  
TCAGTCACTGCAACCTCTGCCTCCCTGGCTTAAGCGATCCTCCACCTCAGCCTCCTGA  
GTACGTGTGACCATAGGCCATGGCACAAAGCCAGCTAATTTTTTTGATTTTTTAGTAGA  
AATGTGGTTTACCATGTTGCATAGGCTGGTCTCGAAGTTCTGAACTCAAGTGATCTGCC  
TGCCTTGGCCTCCCAAGTGTCTGGGATTCTAGGTATGAGCCACCCTGCTCGGCCTATAAT  
GGCACTTTCTATCCCAATTGATGAGGCTCTACTCTCATGACCTAATCATCTCCCAAGGC  
[C, G]  
CTAAGGCCTCCTGATACCATCACCTTTGGGGTTAGGTTTTAATATATACATTTTGGGGG  
ACACAGACATTTAGACCATAGCACCTCCATTGAAAGGAAACATTTCTGACACCTGGCTA  
TCTCAAGGGGCCCTTTCAGTTCCCTGCAGGCTGCATTCACATCAACCAAGAGCAG  
CGACACTCACTCAGAGGTTAAATAACTTGTCCAGAGTCACAGCAGTAATGAATGACAGAG  
CTGGGCTTGAATCCAGGCTCCTCCTAGAGCCTGGATTCTGTGTAGTGAGTGAAAGCTG

21796  
CATTCCACATCAACCAAGAGCAGCAGCACTCACTCAGAGGTTAAATAACTTGTCAG  
AGTCACAGCAGTAATGAATGACAGAGCTGGGGCTTGAATCCAGGCGTCTCCTAGAGCCT  
GGATTCTGTGTAGTGAGTGAAAGCTGACTCCTGGGAGACTTCTGCGTGGTCTGGTCTC  
TCTCCAGACTGCACTGCGCAAGTTTCTCTTCTGATGGTCCCTAGGGTATTACAAAGACA

FIGURE 3, page 19 of 22

GTGGCCCTGCCTGTCTCAGGTGTTTTATTACCAGATGAGGTCATGGCCTCAGGAACCCCTGT  
[A, G]  
GGAAGCTGAGTTTCAGAGTCTTTGAGCAGGCTTTAGGGAGGTTCCAGCTTCCCACCACCAA  
GCCCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGTGGGGGGTCTGGAAGATGAGGTTG  
CGGGGTGCGATATTCTGCCCCAATTCGCCCCCTCCTTGCTCAATCTGTTTCTGCAGGTATTG  
CTGACTACAGACCCCAAGGATGGAGAAACCATTGAGCTGAGGCTGGTTAGCTGGTAGCCCC  
TGAGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTGATG

22159 CCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGTGGGGGGTCTGGAAGATGAGGTTGCG  
GGGTGCGATATTCTGCCCCAATTCGCCCCCTCCTTGCTCAATCTGTTTCTGCAGGTATTGCT  
GACTACAGACCCCAAGGATGGAGAAACCATTGAGCTGAGGCTGGTTAGCTGGTAGCCCCTG  
AGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTGATGTC  
CCTGGAACAGGAACTCGCCTGACCTGCTGCCACTCCTGTGCACTTTGAGCAATGCCCC  
[C, T]  
TGGGATCACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGGAGCAG  
AGAGCCAAGCATCTTCCCTGGGAAGTCTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCA  
GGTCTCCCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGGC  
AAAAACGGAGTCCGAGGCCGAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTC  
CTGCAAGAAGGCCTCCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCT

22334 CCCTGAGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTG  
ATGTCCTTGGAAACAGAACTCGCCTGACCTGCTGCCACCTCCTGTGCACTTTGAGCAAT  
GCCCTTGGGATCACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTG  
GAGCAGAGAGCCAAGCATCTTCCCTGGGAAGTCTTCTGGCCAAGTCTGGCCAGCCTGGC  
CCTGAGGTCCTCCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTC  
[C, A]  
TTGGCAAAAAACGGAGTCCGAGGCCGAGGTGTTGTGAAGACCACTCGTTCTGTGGTTG  
GGGTCTGCAAGAAGGCCTCCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACT  
CTGCTGTTAGAGTGGCAGCTCCGAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCA  
GGGCTGCTCAGTGCCTGCCTCTGACAAAATTAAGCATTGATGGCCTGTGGACCTGTAC  
AGTGGCCTGGTGCCTCATACTCCTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGGAAGTA

22346 TCATCCCAGCAGCCTCGCACACTCCCTAGGCTTCTACCCTCCCTCCTGATGTCCCTGGAA  
CAGGAACCTCGCCTGACCTGCTGCCACCTCCTGTGCACTTTGAGCAATGCCCCCTGGGAT  
CACCCATGGGATCACCCAGCCACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTG  
AAGCATCTTCCCTGGGAAGTCTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCAGGTCTC  
CCATGAAGGCCACCCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGGCAAAAA  
[C, -]  
GGAGTCCGAGGCCGAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTCTGCAAG  
AAGGCCTCCTCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGTGTTAGAG  
TGGCAGCTCCGAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCAGGGCTGCTCAGT  
GCCTGCCTCTGACAAAATTAAGCATTGATGGCCTGTGGACCTGTACAGTGGCCTGGTG  
CCTCATACTCCTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGAAGTAACCCCATCAGGG

22476 ACAAGCCCTTCGAGGGCCCTATACCATGGCCACCTTGGAGCAGAGAGCCAAGCATCTTC  
CCTGGGAAGTCTTTCTGGCCAAGTCTGGCCAGCCTGGCCCTGCAGGTCTCCCATGAAGGC  
CACCCATGGTCTGATGGGCATGAAGCATCTCAGACTCCTTGGCAAAAAACGGAGTCCGC  
AGGCCGAGGTGTTGTGAAGACCACTCGTTCTGTGGTTGGGGTCTGCAAGAAGGCCTCC  
TCAGCCCGGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGTGTTAGAGTGGCAGCTC  
[C, T]  
GAGCTGGTTGTGGCACAGTAGCTGGGGAGACCTCAGCAGGGCTGCTCAGTGCCTGCCTCT  
GACAAAATTAAGCATTGATGGCCTGTGGACCTGTACAGTGGCCTGGTGCCTCATACTC  
CTCAGGTGCAGGGGCAGGGACAAGAGAAGGGGAAGTAACCCCATCAGGGAGGAGTGGAG  
GGTGCCTGAGCCGCATGTGGGCATTGGGGAGTGATGGGAATGCCAGCAGTGATGACGT  
TGACTACTGACTGAGCACCCTACTATGACTGAGCACTCACTCGCTAGATACTATCTTG

23143 GCCGGGCGTGGTGGCTCACGCCCTGTAATCCAGCACTTTAGGAGGCCAAGGCAGGTGGAT  
CACAAGGTGAGGAGTTTGAGATCAGCCTGGCCAACATGGTGAACTCCATCTTTACTAAA  
AATACAAAAAATTAGCCAGGCATGGTGTTCATGCCTGCATGCCTGTAATCCAGTTACT  
TGGGAAGCTGAGGCAGGAGATTGCTTGAACCCCTGGAGGCGGAGGTTGATGAGCCGAG  
ATCACGCCATTGCACTCCAGCTTGGGCAAGAAGAGAAACACTCTCAAAAAAAAAAAAAA  
[A, -, T]  
CAGGAAACTGGTGCTCAAAAAGGAAAAGTGACTCACCAAGGTCACAGACTAGGCAGTGAT  
GCTGGGGGAACCTGGCTCAGGGGACACAGACCTGGCCCTGGGGCAGCCTTGCAGCTCCTCC  
ACTAAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAA  
CTCAACTGGAGCCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTTCCACAAGGCA  
CCATTAGGGCATCACCCAGGCCTCGGCAGCCACGACGAGGATCCTGCCTCTCATTGG

23445 AGGAAACTGGTGCTCAAAAAGGAAAAGTGACTCACCAAGGTCACAGACTAGGCAGTGATG  
CTGGGGGAACCTGGCTCAGGGGACACAGACCTGGCCTGGGGCAGCCTTGCAGCTCCTCCA  
CTAAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAAC

FIGURE 3, page 20 of 22

TCAACTGGAGCCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTTCCACAAGGCAC  
CATTAGGGCATCACCCAGGCCCTCGGCAGCCACGACGAGGATCCTGCCTCTCATTGGT  
[T, C]  
GGGGGCTTAGGGGCTCTGGGCTGCCCTCTTGAAGAGGGGTTACAGCCAGCGAGGCACCC  
CCTATGCTGCACCCACCAAGGTTAGGAAGAGGTCCTGTCTCAGTGGGGCCCTCTGATG  
AACAGCCCATCAGGCTCGGTCCACATGCCCTTGAAGAGATGGTGACATACTCAAAGTCC  
TTGAAGCCGCATATTAAACCACCTAGAGCACCATCTTCAAACATTTAGGGTCTGAGAAGA  
TAGGGGAAGTAAGCAATTTAAACATTTCTTTATATTGGGCCAGGTGCAATGGCTCACGT

23974 GGTCTGAGAAGATAGGGGAAGTAAGCAATTTAAACATTTCTTTATATTGGGCCAGGTGC  
AATGGCTCACGTCTGTAATCCCAGCGCTTTGGGAGGACGAGGATCACCTGAGGTGAGGAG  
TTCAGATCAGCCTGGCCAACATGGAGAAACCCATCTCTACTAAAAATACAAAATTAG  
CTCAGGCGTGGTGATGTGCACCTGTAATCCTAGCTATTAGGAGGCTGAGGCACAAGAA  
TGCTTGAGTCAATATTGCACCACTGCACCTCCAGCCTGGGCAACAGCGAGACTCTTGCTC  
[A, -]  
AAAAAAAAAAAAAGATATTTGCTGAAAAGACCCAGCCTGCCAACTCAGGGGCAGCCAAGG  
GAGGTAGTGAAATGGAAGTTGGAGCTCAGCGCTCCACACCTCCACTGCCCTCAGGCCTT  
CTCTGCCTCTTTCCCATCAGTCAGTGTCTCTGGGCATGGTCTTGGCAGAGACTTGGCCT  
CCTTCCAGTTCAAGCTCCCTCTTAGATTGTGTCCACGCCACTGAGTCTTTGGGACACTG  
GGTCAGATGTCTAGTCTGGCACAATTGGCAGGAATCCAAGAAACAGTGTGAGTGAGGGG

24865 ATAGCTGGGATTACAGGTGTGCACCACCATGCCAGCCTAATTTTGTATTATTAGTAGA  
GATGGGTTTTACCATGTTGTCCAGGCTGGTCATGAACCTCCTGACCTCAAGTGATCCACC  
CGCTTTGGCCTCCCAAAGTGTGGGATTACAAGCATGAGCCACAGTGCCTGGCCTGACCC  
TGCTCTTTTGAAGACCATTCCCCCAAATCTGTGCACCTGTGTGCCTTTCTCTCTCTG  
CCTCCTCTCAGCTCTGCCCCGCTCTCCTCCCTTCTCCTCTGGCAAATCCCACTCATCTCT  
[G, T]  
GAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGGATGAAG  
GACGTGGCCACGGAGTTTGTGTTTGTGTTTGTGAGATGGAGTTTGTCTCATGTTGCCC  
AGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGTTCAAGC  
GGTTCTCCTGCTTAGCCTCCCAAGTGTGGGATTACTGGCATGAACCACCACACCTGG  
CTAATTTTGTGTTTTTAGTAGAGATGGGGTTTTCTCATGTTGGTCAGGCTGGTCTCGAAC

25040 GACCCTGCTCTTTTGAAGACCATTCCCCCAAATCTGTGCACCTGTGTGCCTTTCTTCT  
CTCTGCTCCTCTCAGCTCTGCCCCGCTCTCCTCCCTTCTCCTCTGGCAAATCCCACTCA  
TCTCTTGAAGCCCTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGG  
ATGAAGGACGTGGCCACGGAGTTTGTGTTTGTGTTTGTGAGATGGAGTTTGTCTCATG  
TTGCCCAGGCTGGGTACAAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGT  
[T, C]  
CAAGCGGTTCTCCTGCCTTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCACCACA  
CCTGGCTAATTTTGTGTTTTTAGTAGAGATGGGGTTTCTTCATGTTGGTCAGGCTGGTCT  
CGAATCCCAACCTCAGGTGATCTGCCTGCCTCGGCCTCCCAAAGTACTGGGATTACAGG  
GTTGAGCCACTGTGCCTGGCCACGGCCACGGAGTTTAAAGAGGCTTCTGTGGCAGTGG  
CATCCAGACGGAGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGGAGTG

25046 GCTCTTTTGAAGACCATTCCCCCAAATCTGTGCACCTGTGTGCCTTTCTTCTCTGTC  
CTCCTCTCAGCTCTGCCCCGCTCTCCTCCCTTCTCCTCTGGCAAATCCCACTCATCTCTT  
GAAGCCCTTTCTTCCAGGGGAAGCCCTGATCATGCTGCTTTCTCCTGTGGGAGGGATGAAG  
GACGTGGCCACGGAGTTTGTGTTTGTGTTTGTGAGATGGAGTTTGTCTCATGTTGCCC  
AGGCTGGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCCGGGTTCAAGC  
[G, A]  
GTTCTCTGCCTTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCACCACACCTGGC  
TAATTTTGTGTTTTTAGTAGAGATGGGGTTTCTTCATGTTGGTCAGGCTGGTCTCGAACT  
CCCAACCTCAGGTGATCTGCCTGCCTCGGCCTCCCAAAGTACTGGGATTACAGGTTGAG  
CCACTGTGCCTGGCCAGGCCACGGAGTTTAAAGAGGCTTCTGTGGCAGTGGCATCCA  
GACGGAGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGGAGTGTGAGTT

25193 ATCATGCTGCTTTCTCCTGTGGGAGGGATGAAGGACGTGGCCACGGAGTTTGTGTT  
TTGTTTTGAGATGGAGTTTGTCTCATGTTGCCAGGCTGGGGTACAATGGTACGATCTCA  
GCTCACTGCAACCTCTACGTCCCGGGTTCAAGCGGTTCTCCTGCCTTAGCCTCCCACTCA  
GCTGGGATTACTGGCATGAACCACCACACCTGGCTAATTTTGTGTTTTTAGTAGAGATGG  
GGTTTCTTCATGTTGGTCAGGCTGGTCTCGAATCCCAACCTCAGGTGATCTGCCTGCCT  
[C, T]  
GGCCTCCCAAAGTACTGGGATTACAGGTTGAGCCACTGTGCCTGGCCAGGCCACGGGA  
GTTTTAAGAGGCTTCTGTGGCAGTGGCATCCAGACGGAGTGCAGAACTCAAAGTTGAA  
GGCCAGAAGCTCAGGGAAGGGGGAGTGTGAGTTGAGGAGTCTCTTGCTGCCAGGGCCAG  
AAACCGAAGTCCAAGCCTCTCCACAACAGCGGGTGTAGAGCATGTAGAATCAGAGAGGAG  
GCTGAGCCATGCAGCCCGAGAAGAGGGGAATGCCACTGAGCCACAGAGACCCAGTGCCA

25529 AGTGCAGAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAGGGGGAGTGTGAGTTGAGGA  
GTCTCTTGGCTGCCAGGGGCCAGAAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAG

FIGURE 3, page 21 of 22

AGCATGTAGAATCAGAGAGGAGGCTGAGCCATGCAGCCCCGAGAAGAGGGGAATGCCACT  
GAGCCACAGAGACCCAGTGCCACTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCC  
[T, G]  
GCCTGTCTCTGTATGCAAGGCTTCACCCCTCTCTCGTTGTACATTGTACACATTCTAGGTGA  
CACCAGCAGCTTCTGATTCTCATCTCCCATAACATCAGCCCCCAGAGAGGGGACAACCTG  
CTGAGCTGATAACATAATAGATGCCCTTTCTTGAGGCCATGGTCATGGTCAGCGTGGA  
GAGGATGAAGCCTGAGCAGGCAGGATCGGGGGTCTAGAGGGGAAGGAGGTGGAAGTT

25614 GGGCAGAAGCTCAGGGAAGGGGGAGTGTGAGTTGAGGAGTCTCTTGGCTGCCAGGGCCAG  
AAACCGAACTCCAAGCCTCTCCACAACAGCGGGTGTAGAGCATGTAGAATCAGAGAGGAG  
GCTGAGCCATGCAGCCCCGAGAAGAGGGGAATGCCACTGAGCCACAGAGACCCAGTGCCA  
CTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCCGGCCTGTCTCTGTATGCAAGGCTTCA  
CCCTCTCTCGTTGTACATTGTACACATTCTAGGTGACACCAGCAGCTTCTGATTCTCATC  
[C, T]  
CCCATAACTCAGCCCCCAGAGAGGGGACAACCTGCTGAGCTGATAACATAATAGATGCC  
CCTTTCTTGAGGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGAGCAGGCAGGA  
TCGGGGTCTAGAGGGGAAGGAGGTGGAAGTTGAGATCACAGACCTGTGGTCAGGTGGCC  
TGGGAAGGGTTTGACAGAGTGTGCGCCCAAAGAGCTTGAAGGGATTTTGCTGCTGTGGGT  
GAGCACTGCCTCTCCCTTAGGGACAACAGCCACCTCTTCTCTCCCATTTGCTTTTCCC

25785 CCAGTGCCACTGCCAGGTGTCTCTGCCTCCACTTCCCATGACCCGGCCTGTCTCTGTATG  
CAGGCTTCACCTCTCTCGTTGTACATTGTACACATTCTAGGTGACACCAGCAGCTTCTG  
ATTCTCATCTCCCATACATCAGCCCCCAGAGAGGGGACAACCTGCTGAGCTGATAACAT  
AATAGATGCCCCCTTTCTTGAGGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGA  
GCAGGCAGGATCGGGGGTCTAGAGGGGAAGGAGGTGGAAGTTGAGATCACAGACCTGTGG  
[C, T]  
CAGGTGGCCTGGGAAGGGTTTGACGAGTGTGCGCCCAAAGAGCTTGAAGGGATTTTGCT  
GCTGTGGGTGAGCACTGCCTCTCCCTTAGGGACAACAGCCACCTCTTCTCTCCCATTT  
GCCTTTCCCTTCTGTAGATATGAAACACAGGCCTCCTGTGAGGCCCTACTTAACCTCC  
GTGATGGGGAAGCGGCCGGAGAAAGGGAGTTCTGGCAGCTTCTCCGAGACCCCAACACC  
CCACTGTTGCAAGGTGAGTCATGGCCTGACACTCTGGATGTGTCCCTACCCCAAGCTTA

26266 GTGATGGGGAAGCGGCCGGAGAAAGGGAGTTCTGGCAGCTTCTCCGAGACCCCAACACC  
CCACTGTTGCAAGGTGAGTCATGGCCTGACACTCTGGATGTGTCCCTACCCCAAGCTTA  
CTCAGCCAAGAGGCTTCATCAACTACCCCAAGCTTTCCCTAGCACCTCCTGGGCCACAC  
CTTCACAAAATCACTGATGCTCAAAGTTGGATATAATATATTGAAGCTGAAGCCTTAGCAT  
TTTTATGCAAGTTACTGTGGAATTCTAGGAAACCAGACAGATTACAAAAA  
[A, -]  
CTAGAAGAAAATTAACATCACCTAGGATATACTACCTAGGAATAACGTCTTTTATTTGA  
GATGGAGTTTCGCTCTTGTTGCCAGGCTGGAGTGCAGCGGTATGATCTCGGCTCGCTGC  
AACCTCCGCCTCCTGGGTTTATGTGATTCTTCCACCTCGGCCTTCTAGAGCCCAAGTGG  
TCGCTGCTCTGCTCCCAAGTTCTGGGATTACAGGCATGAGCCACCGCACCCAGCC  
AAAATTACTTAACCTTTCTTCTAGATACTTTTAAAAATATGGCAGTAAGTTTTTCATAA